December 2020

The SHAPE of U: Mapping Out Protective Elements in mRNA Escapees

Jacob Miles

The SHAPE of U: Mapping Out Protective Elements in mRNA Escapees

A Thesis Presented

By

Jacob Matthew Miles

Submitted to the Graduate School of the University of Massachusetts Amherst in partial fulfillment of the requirements for the degree of

Master of Science

September 2020

Microbiology Graduate Program

The SHAPE of U: Mapping Out Protective Elements in mRNA Escapees

A Thesis Presented

By

Jacob Matthew Miles

Approved as to style and content by:

_____

Mandy Muller, Committee Chair

_____

Peter Chien, Committee Member

_____

Michele Klingbeil, Committee Member

_____

James Holden, Department Head

Graduate Program in Microbiology

ABSTRACT

THE SHAPE OF U: MAPPING OUT PROTECTIVE ELEMENTS IN

MRNA ESCAPEES

SEPTEMBER 2020

JACOB MATTHEW MILES, B.S. UNIVERSITY OF NEW HAVEN

M.S., UNIVERSITY OF MASSACHUSETTS AMHERST

Directed by: Professor Mandy Muller

A crucial step of the viral life cycle of Kaposi's Sarcoma Herpesvirus (KSHV) lytic infection is the triggering of a massive RNA decay event termed "Host Shutoff". Host Shutoff is driven by the viral endonuclease SOX which leads to the destruction of over 70% of the total transcriptome. This process cripples cellular gene expression and allows for viral reprograming of the cell for the purpose of viral replication. Co-evolution has led to the host developing a multitude of antiviral defenses aimed at preserving certain cellular RNAs linked to antiviral responses. One such defense are RNA secondary structures located within the 3'UTR of select host transcripts that protect them from SOX degradation. This structure, known as the SOX Resistant Element or SRE, has previously been isolated to a 200-nucleotide region found within the 3'UTR of the host transcript Interleukin-6. In this thesis, I sought to further define the structure of the IL-6 and other SREs using SHAPE-MaP to generate chemically-probed RNA structural models. Through this work, I demonstrated that the IL-6 SRE confers a form of active resistance to SOX cleavage, and based on structural analyses,

likely acts as a scaffold for the recruitment of a protective ribonucleoprotein complex. This research highlights the importance of RNA secondary structures in influencing mRNA fate during viral infection and establishes the groundwork for understanding how these structural features can facilitate escape of cellular transcripts from viral endonucleases.

TABLE OF CONTENTS

Page

LIST OF TABLES

Table Page

LIST OF FIGURES

Figure Page

CHAPTER 1

INTRODUCTION

1.1 Kaposi's Sarcoma-associated Herpesvirus, Host Shutoff, and Escape

Kaposi's Sarcoma-associated Herpesvirus (KSHV), **Figure 1** also known as human herpesvirus 8 (HHV-8), is a member of the gammaherpesvirus family, a group of double-stranded DNA viruses that are responsible for the establishment of lifelong viral infections. KSHV is the causative agent of Kaposi's sarcoma, a cancer that causes neoplasms, in the form of lesions, on the skin, oral cavity, and major organ systems [Ganem 2006]. The condition can be sorted into four predominant forms:

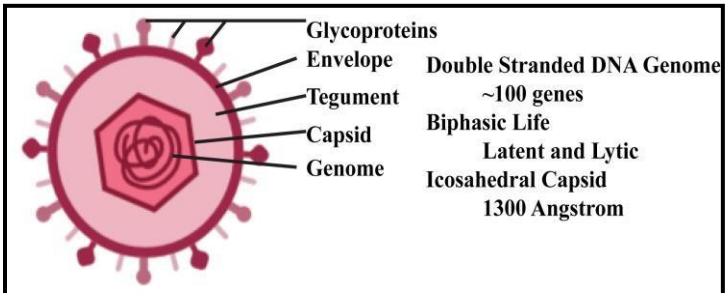

Figure 1: KSHV Representation: Representation of KSHV Virion with labels

- Classical KS- Endemic to the Mediterranean region and characterized by minor, clinically insignificant, skin growths in elderly males
- Endemic KS- Endemic to Africa, targets young children and affects lymph and organ systems [Mesri, Cesarman, and Boshoff 2010]
- Iatrogenic KS- Associated with individuals immunocompromised from organ transplant, highly clinically significant, leads to cancer and death
- AIDS associated KS- Associated with individuals immunocompromised from AIDs, highly clinically significant, leads to cancer and death [Ganem 2006, Mesri, Cesarman, and Boshoff 2010].

KSHV targets a diverse range of host cells from endothelial to epithelial and B lymphocytes [Yan et. al. 2019]. Maintaining a prolonged infection requires stringent control of

the host defense systems. To address this severe anti-viral state, KSHV promotes a biphasic lifestyle, riding a careful balance between gene expression and control of the host immune sensors. Infections are established in a latent phase in which there is minimal viral gene expression, and the viral genome is restricted to the nucleus as an episome with no genomic replication or virion production [Ganem 2006]. Under certain conditions of cellular stress, such as host immune sensing of viral activity or oxidative stress, KSHV infection switches to a lytic phase [Brulois and Jung 2014]. This event triggers the upregulation of viral gene expression leading to the replication of the viral genome and the production of viral progeny [Jenner et. al. 2001]. This acts as a last resort escape mechanism where the virus increases viral gene expression and attempts to usurp host gene expression and cellular functions. This ensures that KSHV is able to replicate and spread while protecting other infected cells from host detection by targeting interferon production and signaling [Burýšek et. al. 1999, Zhu and Yuan 2003].

During this lytic phase, where viral activity is at its highest, the virus suppresses further host immune activity pathways, as well as co-opting others, to prevent cellular apoptosis and promote replication [Gwack et. al. 2001, Johnston, Pringle, and McCormick 2019, Zhang, Ni, and Damania 2020, Tabtieng, Degterev, and Gaglia 2018]. To optimally handle all these pathways, KSHV triggers a major event known as Host Shutoff that occurs during the early phases of lytic reactivation. Host

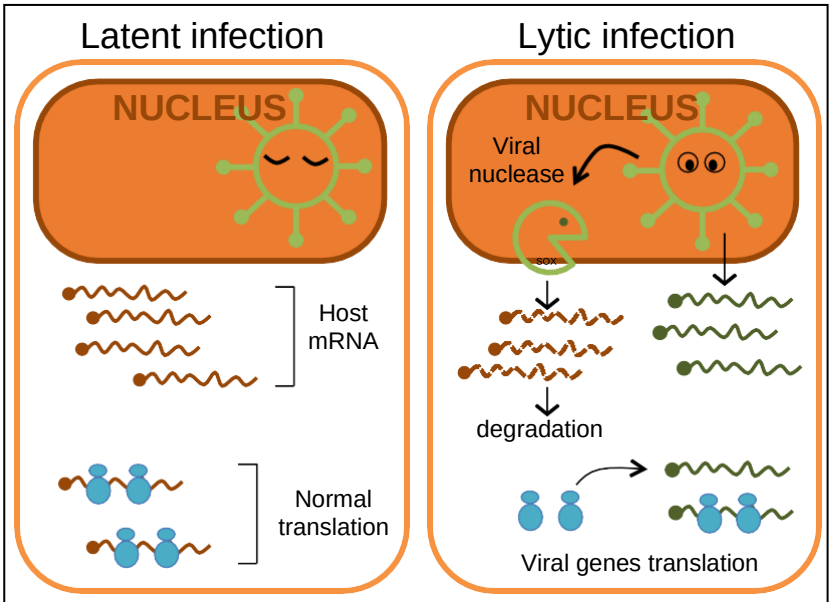

Figure 2: Viral Expression of SOX: During lytic infection, herpesviruses usurp the host translation machinery by using a viral nuclease that degrades most of the host mRNAs.

shutoff is characterized by the widespread degradation of a large fraction of cellular transcripts leading to a significant downregulation of host protein production. This process shifts cellular machinery towards viral needs as well as stopping host immune responses. Host shutoff can be driven by a diverse range of mechanisms, such as blocking nuclear export, direct degradation of host transcripts, disruption of host transcript stability by 5'cap removal or by interfering with translation [Narayanan, Makino 2013, Rivas, Schmaling, and Gaglia 2016, Rodriguez et. al. 2020]. In gammaherpesviruses, the primary method for coordinating host shutoff is mRNA transcript degradation by viral endonucleases. To achieve host shutoff, these endonucleases broadly bind to target sequences within mRNA. After binding to the endonuclease catalytic site, the transcript is cleaved and the remaining fragments are degraded by the host RNA decay machinery [Gaglia, Rycroft, and Glaunsinger 2015]. In KSHV this endonuclease is a protein known as Shutoff Alkaline Exonuclease, **SOX (FIGURE 2)**. SOX belongs to a highly conserved family of nucleases known as the PD-(D/E)XK nuclease family and shares a highly conserved structure and sequence with its gammaherpesvirus homologs BGLF5 from Epstein Barr Virus and muSOX from Murine Herpesvirus 68 [Rivas, Schmaling, and Gaglia 2016]. It was described that these endonucleases, including SOX, target RNA polymerase II transcribed mRNAs that are ready for translation, but have not yet associated fully with translational machinery such as the 40s ribosomal subunit [Gaglia et. al. 2012]. The effect of these endonucleases is wide-spread as it is estimated that 80% of total cellular mRNA are rapidly degraded upon expression. The virus is then able to hijack host machinery to increase production of viral proteins and generate new virions. [Lee et. al. 2017, Chandriani and Ganem 2007].

Degradation of host mRNA has extensive consequences for both the infecting virus and the host, however little is known about transcripts that escape this decay event. Do these transcripts favor the host or do they favor the virus? Is escape some evolutionary mechanism designed to protect from viral nucleases, or is it due to viral selection to preserve transcripts that

benefit viral functions? Of the mRNA that escape SOX based decay there is a dichotomy: certain transcripts are seemingly spared from SOX induced degradation at their basal level of expression [Clyde and Glaunsinger 2011]. This seems to suggest that during viral infection these transcripts are not normally targeted and remain at steady state levels in the presence of SOX most likely due to the absence of a SOX targeting motif [Clyde and Glaunsinger 2011, Gaglia, Rycroft, and Glaunsinger 2015, Muller and Glaunsinger 2017]. This is referred to as passive escape as there is no internal mechanism that protects, but only a lack of targeting by the viral nuclease. On the other hand, some transcripts like Interleukin-6 are directly refractory to SOX despite containing a robust SOX targeting element, suggesting some form of dominant, or active escape mechanism [Glaunsinger and Ganem 2004 2, Hutin, Lee, and Glaunsinger 2013].

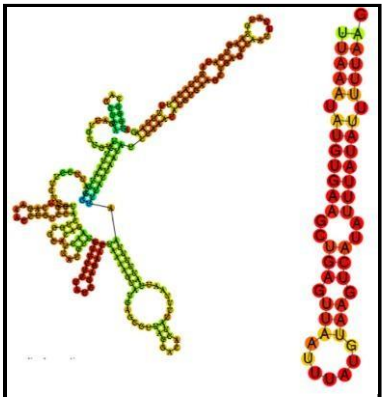

Figure 3: IL-6 SRE Model: Predicted model of the Interleukin-6 SRE with the major hairpin loop enhanced. Mutations that disrupt the hairpin loop abolish SOX resistance. Adapted from Muller and Glaunsinger 2017

Very little is known about these "true" escapees, as most of our knowledge comes from the study of IL-6. Interleukin-6 escape has been shown to depend on the 3' untranslated region (3'UTR), a region of the transcript that is commonly associated with regulation of stability, localization, and translation of a given mRNA [Hutin, Lee, and Glaunsinger 2013]. Contained within this region is an RNA element that has been dubbed the **SOX Resistant Element, SRE**, a 200 nucleotide region that is sufficient to confer protection from SOX decay. Intriguingly, the ability of the SRE to protect IL-6 is transferable to other mRNA, even those that normally would be degraded by SOX. This element was shown to act as a scaffold for a Ribonucleoprotein complex composed of several RNA binding proteins [Muller et. al. 2015]. This region was shown to be necessary to escape from the SOX endonuclease nucleases and was predicted to adopt a complex secondary structure [Muller and Glaunsinger 2017]. Mutational work showed that disruption of this RNA fold is

enough to render IL-6 sensitive to SOX-induced decay. This suggests that this RNA structure is of significant importance to the function of the SRE, and may be the defining characteristic of the resistance, possibly by acting as a binding platform for the ribonucleoproteins. Since the initial discovery of IL-6 resistance to SOX, two other escaping transcripts have been identified, GADD45B and C19ORF66 [Muller and Glaunsinger 2017, Rodriguez, Srivastav, and Muller 2019]. Similarly to IL-6, GADD45B has been shown to contain a SRE-like element within its 3'UTR, with a predicted predominant stem-loop (**Figure 3**). However, the mechanism of escape for C19ORF66 has not yet been identified, though structural predictions of its 3'UTR have discovered several stem-loop motifs that are structurally similar to those contained within the IL-6 and GADD45B.

At this time there are many uncharted mysteries surrounding the SRE. Is it a highly conserved structural element that has evolved in response to viral endonucleases to preserve vital antiviral transcripts? How many transcripts contain this element or elements similar to it? How does this element truly work to prevent SOX binding and/or activity? Since the structure has been shown to be necessary to function, does it require specific protein binding for resistance or is the structure itself mediating protection? Additionally at this time, the true structure is not known; only predictive models exist. My thesis work has been focused on getting us closer to answering these long-standing questions. By exploring the structure of the SRE using structural approaches based on chemical probing, my hope is to ascertain a definitive structure for this remarkable "RNA decay resistant element". Through the lens of my structural work we hope to uncover glimpses of the true mechanisms of the SRE. As form follows function, this RNA structure must be vital to the function of this resistance phenotype, and by going down the RNA structure rabbit hole, our hope is that we will be able to establish a clear link between the structure of the SRE and its function.

CHAPTER 2

BACKGROUND

2.1 RNA Structure

2.1.1 RNA Structural Biology: An Introduction to Structures and Their Dynamics

As this thesis heavily focuses on the investigation of the structure of the SRE RNA element, this section will focus on guiding the reader through some of the basics of RNA structure and folding. Traditional views often depict RNA as a single, spiraling strand that is simple and whose sole role is to code for proteins. But what about the highly structured tRNAs that identify specific coding sequences and chaperone amino acids? What about the rRNA that catalyzes the joining of amino acids to form primary structures? These molecules, along with the discovery that RNA like ribozymes are capable of catalyzing biological reaction, prove that RNA is more complex that initially thought. This complexity is tightly linked to a range of complex structures and folds that these RNA can adopt. Even mRNA, while often single stranded, have been revealed over the past decade of investigation to rarely be linear. The structure of an mRNA is a crucial element that not only brings stability to the molecule, but also contributes to an extensive amount of post-transcriptional interactions that are essential to a multitude of biological processes [Chastain and Tinoco, Jr 1991]. Stability through structure often stems from interactions between Watson-Crick base pairing and internal coaxial stacking, providing interactions that help maintain hydrogen bonding as driven by hydrophobic interactions [Doudna, Doherty 1997]. Much like proteins, RNA structure can be categorized into groups based on the level of complexity and interactions. Primary structure is purely the sequence of the RNA, much as the primary structure of a protein is its amino acid sequence. Secondary sequence in RNA refers to the singular interactions between Watson-Crick base pairing or the lack thereof, such as

the formation of bulges, junctions or loops. Finally, tertiary structure, or motifs, represents interactions between secondary structures within the molecule that make up the three-dimensional structure, with the pseudoknot being one of the most well known structural motifs, and the clover of tRNA being one of the most well known tertiary structures [Chastain and Tinoco, Jr 1991, Batey, Rambo, and Doudna 1999].

RNA structure is driven heavily by complex base pair interactions directed by both hydrogen bonding and π-π stacking interactions of the bases as well as base pairing orientation [Halder and Bhattacharyya 2013]. The canonical base pairing of RNA; A:U and G:C, are known as cis orientation Watson-Crick/Watson-Crick base pair. However, this canonical base pairing, while important for the structural stability in RNA, is not the strongest factor to RNA secondary structures. In fact, non-canonical base pairing, especially that of Watson-Crick mismatches, often known as a Wobble base pair, are the driving forces of the generation of secondary structure. The most significant Watson-Crick mismatch is the G-U wobble pair. This base pairing has increased thermodynamic stability compared to other mismatch groupings, which causes G-U base pairs to be frequently present in secondary structures that split double strands or are involved in sharp turns, because the G-U pair will help stabilize backbone turns [Varani and McClain 2000]. However, not all mismatches are able to form base pairs and instead form extruding elements that are often the basis of several different secondary structures such as loops, hairpins and bulges. These secondary structures are biologically significant in that they not only play a role in the stability of a transcript, but also often have functions that are vital to normal cellular functions or are involved in functional RNAs such as ribosomal RNA or transfer RNA. In the following section we will explore some of the most important secondary structures, highlighting their structural significance and the interactions that drive their form and function.

2.1.2 Types of Secondary Structures

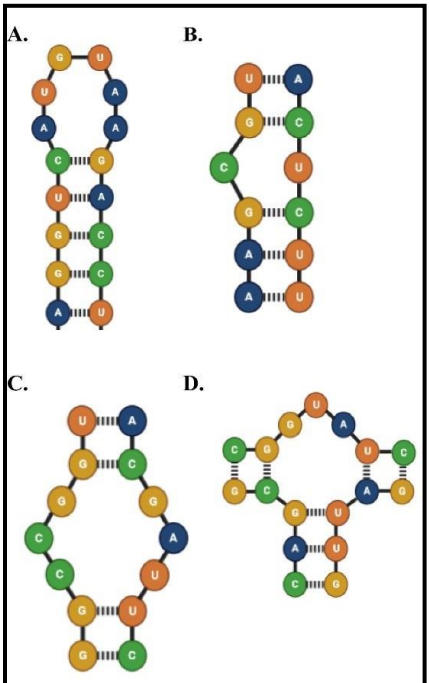

Figure 4: RNA Secondary Structures: Structural Representation of : **A**. RNA Hairpin, **B**. Bulge **C**. Internal Loop, **D**. Multibranch Junction

- RNA hairpins: One of the more biologically significant structures that has been associated with a wide range of biological activities, including regulation of gene expression, is the RNA hairpin, also known as a stem-loop. This structure is the most commonly detected RNA secondary structure, though the structures of hairpins vary drastically from one RNA to the next [Svoboda and Cara 2006]. The structural basis for an RNA hairpin involves a double stranded region of RNA that ends in an open loop [**Figure 4A**]. The variation found in this structure is based on several factors. Loop variety is based on the number of nucleotides involved, where evidence supports loops as small as two unpaired nucleotides, but the most stable structures contain either four or five nucleotides in the loop [Chastain and Tinoco, Jr 1991]. Further variation comes from the presence or absence of internal bulges, the size of these bulges, and the number of bulges within the structure [Svoboda and Cara 2006]. These factors, along with overall length of the structure, leads to significant diversity in the classification of RNA hairpins. This degree of structural versatility is essential to the biological functions that have been associated with these structures. As RNA hairpins are able to exert function in two well defined manners; as cis-acting elements where the sequence and internal structure act as regulatory elements or as a binding platform for trans-acting factors which function through interactions with the RNA structural elements.

- Bulges and internal loops: the bulge structural element occurs when a duplex, or double stranded region is interrupted by unpaired nucleotides that just out from the strand [Wyatt, Puglisi, and Tinoco Jr. 1989]. This structure can be composed of a singular nucleotide that does

not participate in duplexing, or multiple nucleotides that bulge out from the strand with a fair degree of flexibility and can form a number of conformations [**Figure 4B**][Wyatt and Tinoco Jr 1993, Hermann and Patel 2000]. These conformations may include the base remaining within the helix and participating in base stacking with adjacent base pairs. With enough nucleotides participating in the bulge, we can also observe the formation of a kink within the helix leading to a slight turn in the structure, known as a *kink turn* which often changes the direction of linear sections [Huang and Lilly 2016]. One major structural function found within this categorization is the formation of the bulge "flap" in which the bulge can cover a ligand, such as a peptide or small molecule [Hermann and Patel 2000].

Internal loops, much like bulges, depend on mismatched nucleotides, where a non-Watson-Crick parings lead to structure elements. The smallest of the internal loops is the mismatch, occurring due to a single non-Watson-Crick pairing leading to a small loop [**Figure 4C**][Wyatt and Tinoco Jr 1993]. As these structures grow in size, G·U wobble pairs often play a role due to their shortened backbone angles, which can often lead to decreased stability of G·U pairs in certain duplex structures and the breaking of hydrogen bonds to form mismatches [Varani and McClain 2000, Chastain and Tinoco Jr 1991]. The uniqueness of the G·U wobble pair means that it can play a stabilizing role in junctions forming tight bonds and turns as well as being able to participate in mismatch regions with weaker bonding interactions, forming complex structures within RNA.

- Multibranch loops: The final secondary structure that will be discussed is the junction or multibranched loop. These structures represent a region where three or more duplex regions are joined together at a junction of unpaired nucleotides [**Figure 4D**][Chastain and Tinoco, Jr 1991]. This structure is best observed in tRNA where each of the stems comes off of a four-stem junction. The importance of junctions is that it allows for the participating helical regions to take part in coaxial stacking, promoting greater thermodynamic stability [Wyatt, Puglisi, and Tinoco

Jr 1989]. In three helix junctions some trends have been reported including selection for right-handedness in the stacking of bases and the alignment of the backbone as well as more internal interactions within the minor groove as apposed to the major groove [Lescoute and Westhof 2006]. While three-way and four way junctions are the most frequently observed, up to 9-way junctions have been observed [Binderwald et. al. 2008].

2.1.3 Types of Tertiary Structures

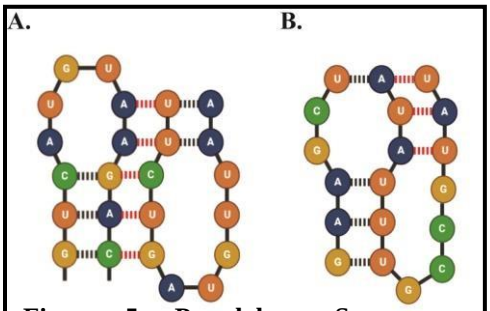

Figure 5: Pseudoknot Structures: Representation of two different forms of a pseudoknot: **A**. Pseudoknot involving two hairpin loops folding together **B**. Pseudoknot involving single stranded RNA folding over the hairpin loop forming interactions

Tertiary structures represent places where interactions occur between secondary structures within the RNA element, often times, when an RNA molecule is represented as a 3D element this is a representation of the tertiary structures and is define by tertiary motifs. Some of earliest demonstrated tertiary interactions were those observed in tRNA, such as the coaxial stacking interactions between the different RNA stems of the clover [Batey, Rambo, and Doudna 1999]. The easiest way to classify tertiary structures is to group them into motifs based off of the type of internal interactions they participate in, whether this is helix-helix interactions, helix-non-helix interactions, or interactions that involve two non-helix regions that help stabilize three-dimensional structures. One of the largest barriers to the study of RNA tertiary structure is generating three-dimensional data, as one needs to generate structural data through experiments such as X-ray crystallography, nuclear magnetic resonance (NMR), cryo-EM, or other structural methods to generate data that can be interpreted into 3D interactions [Chastain and Tinoco, Jr 1991; Westhof and Auffinger 2000; Magnus et. al. 2019]. Due to this, most information on tertiary RNA structures are based off of the most abundant RNA molecules such as tRNA, the Hammerhead ribozyme, the P4-P6 domain of group 1 introns, rRNA subunits, riboswitches and other such structures that have been

thoroughly characterized [Chastain and Tinoco Jr. 1991, Westhof and Auffinger 2000, Abraham et. al. 2008]. Because of these limitations, a lot of tertiary structure analyses are inferred from phylogenetic approaches, such as comparative sequence analysis, or predictive computer modeling based on combinations of theoretical data such as sequence alignment, experimental data, and thermodynamic calculations such as base pair constrains and interactions as well as the classification and identification of motifs [Westhof and Auffinger 2000, Magnus et. al. 2019].

- Pseudoknots: Many of these tertiary motifs and interactions represent the form in which biologically significant interactions take place, with some interactions being associated with RNA stability or resistance to endonucleases. Of all the tertiary interactions the most well known and understood is the pseudoknot **Figure 5**. This structure is often lumped in with secondary structure because it is easy to predict and can be represented in a two-dimensional plane. However, because it captures the interaction of two helical segments connected by single-stranded regions or loops as well as the interactions of two hairpin loops to partake in coaxial stacking, it is a tertiary interaction [Staple and Butcher 2005, Batey, Rambo, and Doudna 1999]. One of the best examples of a pseudoknot for this thesis is one found within the 3'UTR of the genome of flaviviruses [Steckelberg, Vicens, and Kieft 2018]. Within this region there is a pseudoknot structure that forms a ring-like conformation which is able to stall the endonuclease function of the host XRN1 exonuclease. This structure has been identified in all the flaviviruses and acts as an exoribonuclease-resistant element which protects viral transcripts from host decay functions [Steckelberg, Vicens, and Kieft 2018, Ochsenreiter, Hofacker, and Wolfinger 2019]. This incredible level of fine-tuning RNA structures to counteract host defenses therefore emerges as a common theme in the virus world that we will further explore in this thesis. It also exemplifies how evolution can shape the regulation of viral expression by precisely manipulating RNA fold and thus highlight the importance of better exploring these complex questions.

- Tetraloops: These structures are often found within hairpin loops which form tertiary interactions with structures in other regions of the RNA known as tetraloop-receptors and lead to molecular stabilization [Batey, Rambo, and Doudna 1999, Wyatt and Tinoco Jr 1993, Butcher and Pyle 2011]. Another major tertiary interaction is that of the triple helix which forms when a single strand region which interacts with a double stranded region [Chastain and Tinoco Jr 1991, Batey, Rambo, and Doudna 1999, Butcher and Pyle 2011]. One triple helix of note is an element known as the expression and nuclear retention element (ENE) of the KSHV PAN RNA, a long noncoding RNA [Rossetto and Pari 2014]. In this structure the polyA tail of the RNA folds into a prominent stem-loop structure and wraps into the structure forming a triple helix that protects the RNA from deadenylation and decay [Mitton-Fry et. al. 2010]. This once again highlights how viruses are evolutionarily compelled to counteract to the host defenses and to co-opt the complex world of RNA structure for their own advantage.

This brief overview shows that RNA is a complex molecule, able to form a wide arrange of biologically significant structures and interactions. Many of these structures are widely prevalent, but the presence of biologically significant structures in mRNA is beginning to been seen as significant in the scientific limelight. This is in part due to the development of new techniques and methodologies. Previously the only methodology for accurately studying RNA structural elements was through time-consuming techniques such as X-ray crystallography, which also demanded highly detailed understanding of the data output. But the era of high-throughput transcriptomics, chemical probing, and modeling has led to many novel structural discoveries in the field of RNA biology. As such, mRNAs that once were ignored structurally can now be explored with relative ease, and novel RNA structural elements are being discovered on a regular basis. Understanding these elements will unveil new interactions, not just amongst host cells, but also in the realm of pathogen interactions as we are able to better understand the mechanisms through which viruses target and exploit host genetic material.

2.2 mRNA Biology: Exploring Structure, Interactions, and Cis- and Trans- Acting Elements

2.2.1 An Introduction to mRNA Regulatory Elements

Modern technologies have brought more attention to noncoding regions of mRNA, specifically the 5' and 3' untranslated region (UTR). The 5'UTR is upstream of the coding region following directly after the 5'cap while the 3'UTR is downstream of the coding region. Each region regulates separate mechanisms important for the mRNA transcript life cycle, ranging from translational control. to the regulation of stability and transport [Mignone et. al. 2002]. While variation exists within some regions of the 5'UTR there is generally more conservation within specific 5'UTR regions. This is due to the 5'UTR playing a vital role in the recruitment of the initiation complex which is dependent on conserved structural and sequential elements [Pesole et. al. 2001, Mignone et. al. 2002]. Comparatively, 3'UTRs are highly diverse with many different structures and sequence elements, influencing a range of 3'UTR-associated functions. While this thesis focuses heavily upon the 3'UTR and its interactions, this section will briefly cover the 5'UTR and explore its relation to the 3'UTR.

While both the 5' and 3' UTRs play major roles in the lifecycle of a transcript, they bare significant structural and functional differences. One of the most striking differences between the structure of the 5' and 3' UTRs, is relative size. 5'UTRs tend to average out to only 200 nucleotides in length in humans with the longest known in humans being roughly 2800nt in length [Leppek, Das, and Barna 2018, Mignone et. al. 2002]. In contrast, 3'UTRs have an average length of around 1000nt with the largest being over 8000nt long. It is believe that the relative conservation of short 5'UTR is due to the importance of this region in regulating translation initiation and many have shown that the additional sequences or motifs in this region results in

lower translation efficiency [Liu et. al. 2012]. Meanwhile, 3'UTRs are evolutionarily more flexible and have seen a number of translocation events that have resulted in the addition of elements regulating transcript longevity. 5'UTRs also have significantly higher average GC content, ranging around an average of 60% versus the 3'UTR which has an average of 45% GC content [Pesole et. al. 2001, Mignone et. al. 2002]. High GC content acts as a translation efficiency sensor, where higher GC content causes a decrease in translation due to decreased scanning efficiency by ribosomes [Babendure et. al. 2006, Leppek, Das, and Barna 2018, Araujo et. al. 2012]. To combat this scanning efficiency detriment, effector proteins are often recruited to 5'UTR and alter ribosomal affinity to mitigate the deficit of high GC content. However, one constant between 5' and 3'UTRs is that they contain RNA secondary structure, which is further strengthening how important RNA structures are to mediating function.

2.2.2 Exploring the 5'UTR

5'UTRs encompass many highly conserved RNA secondary structures crucial to control post-transcriptional functions. Amongst these, the hairpin structures discussed in the previous section are thermodynamically favorable, requiring no input energy to form, and are capable of effectively inhibiting translation when placed near the 5'cap [Araujo et. al. 2012, Babendure et. al. 2006]. For example, this has been extensively observed in iron regulatory elements, IRE, which are stem-loop based structures found in transcripts for proteins involved in iron storage and transport. In low iron conditions, this structure forms a complex with iron-regulatory proteins which actively block ribosome scanning. Another feature unique to the 5'UTR is the presence of internal ribosome entry points, or IRESs, which are primarily in viral genome and transcripts, though a few examples have been discovered in human transcripts. In viruses these highly structured elements allow for hijacking of ribosomes without the need for a 5'cap thus bypassing the rate-limiting step of translation initiation [Leppek, Das, and Barna 2018]. Cellular IRESs are less structured and depend on RNA binding proteins to assist in their functions. These proteins

are thought to act as RNA chaperones, binding to a specific RNA element in the IRES, usually a stem-loop, and remodeling it and opening it up for ribosome recruitment.

The structural complexity of 5'UTR act as a major hurdle towards the recruitment of translational machinery, requiring the recruitment of several accessory proteins to overcome this. In slowing down translational initiation, the complexity of the 5'UTR allows for proper elongation during translation. To initiate translation, ribosomal scanning requires the recruitment of a strong helicase that can breakdown the structural elements until a stop codon is reached [Araujo et. al. 2012, Leppek, Das, and Barna 2018]. While the 5'UTR is structurally complex, this complexity is mostly important to the process of initiation of translation, and once translation is initiated, these 5'UTR elements are no longer needed and can be unwound. On the contrary, the structural elements found within the 3'UTR are spared from ribosomal scanning and therefore persist long after translation initiation, and can regulate many aspects of RNA fate and function, much more so than the 5'UTR.

2.3 Elements and Interactions of the 3'UTR

The 3'UTR has been called the "regulator of RNA fate" as it can influence translation, localization, and even transcript stability [Mayr 2017]. These functions are regulated by a wide array of *cis*- and *trans*- acting factors. However, only in recent years has our understanding of the importance of the 3'UTR progressed significantly. This is due in part to the complexity of studying regulator elements with secondary structures in tandem with the abundance of alternatively cleaved mRNA with different 3'UTR isoforms, multiple polyadenylation sites, and other post transcriptional modifications [Lianoglou et. al. 2013, Mayr 2019]

2.3.1 Sequence Elements

- AU-Rich Elements One of the earliest discovered motifs of importance and the most common regulatory elements found in 3'UTRs is that on the adenylate uridylate rich elements, AU-rich elements (AREs) [Mayr 2019, Matoulkova et. al. 2012]. These elements were first observed in the cellular c-Fos gene through observations that the viral c-Fos gene acted as a oncogene, able to transform cellular fibroblasts into cancerous cells, while the cellular c-Fos gene was unable to do so [Mayr 2019, Chen and Shyu 1995]. Upon deletion of the 3'UTR of c-Fos, cellular transcripts gained similar oncogenic properties to those observed in the viral transcripts. This suggested that an element within the 3'UTR prevented cellular c-Fos from becoming oncogenic. Further investigation of the c-Fos 3'UTR led to the identification of the first AU-rich element, an element that has since been associated with the ability to destabilize mRNAs.

AU-rich elements are often found within the 3'UTR of transcripts that have short half-lives such as cytokines and oncogenes [Shaw and Kamen 1986, Caput et al. 1986, Mayr 2019]. Recently, it has been discovered that trans-acting factors bind AU-rich elements and confer different effects upon the transcript [Mayr 2019, Barreau, Paillard, and Osborne 2012]. For example, AUF1 binding was originally identified as a marker for mRNA degradation, however, more recently some mRNA were shown to be stabilized by AUF1 recruitment. This variability seems to be cell type-specific, however, there is currently no clear methodology of determining whether AUF1 binding will stabilize or mark a transcript for degradation [Barreau, Paillard, and Osborne 2012]. The Hu family of proteins, specifically HuR, also actively bind AREs, acting specifically as a means of stabilizing the mRNA. A third known AU binding protein is tristetraprolin, TTP, which has been associated with deadenylation and degradation rates for target transcripts. Despite the overwhelming history of AU-rich elements being responsible for regulation of transcript turnover, recent research indicates that ARE-binding proteins may play an even more diverse role in transcript life, modulating processing, transport, and translation beyond

the initial mechanism of transcript stability [Otsuka et. al. 2019]. With 5-8% of all human transcripts containing conserved AU-rich element motifs, it is not shocking that new mechanisms and functions are being discovered for ARE binding proteins. With functions ranging from interactions in splicing and the formation of different isoforms, translational repression, and even viral interactions within the cell, it seems that there is still much more to learn from these elements.

- GU-Rich Elements Another major motif in the 3'UTR is the of the GU-rich element, GRE, which bares a strikingly similar motif of repeated units of the "GUUUG" repeat of the AREs. However, unlike the ARE, these repeats always occur in overlaps and occur 2 to 5 times [Matoulkova et. al. 2012]. To further the similarities to AU-rich elements, GU-rich elements are also highly specific to the 3'UTR and are also involved in transcript stability, decay, and half-life. One potential major difference that has been observed was by Halees *et. al.* who observed that synthetic GRES induce destabilization of the target transcript, as expected, but also led to an upregulation of target protein expression. The authors suggest that this might be due to the fact that the GRE binding protein CUG-BP1 is able to enhance translation while the protein CUG-BP2 competes for the same binding site, but shuts off translation [Halees et. al. 2011]. Much like AREs, GRE function is directed by the interactions with trans-acting elements. The binding proteins associated with GU-rich elements, such as CELF1, and the previously mentioned CUG-BP1 and CUG-BP2, are all related to destabilization of transcripts [Halees et. al. 2011, Vlsova-St. Louis and Bohjanen 2011]. However, CELF1 has also been shown to have functions in pre-mRNA splicing suggesting that function of the GU-rich element is also variable based on the binding protein landscape as well as cellular environment.

2.3.2 Structural Elements

Beyond sequence based cis-acting elements, there are also structurally based cis-acting elements within 3'UTRs. Defining these elements into groups is a difficult task as these elements

are usually highly specific structures that bind target proteins with high affinity. Modern chemical probing methodologies have made studying these structures easier and led to a greater understanding of structural cis-acting elements. One of the biggest struggles of understanding structurally based cis-acting elements is the fact that many of these structures can be volatile, with only a few specific nucleotides within the structure participating in protein binding. Knowing whether the structure itself, a specific sequence, or a sub-element within the structure is the primary interaction point is critical for understanding these interactions. To date, there are few well-defined conserved structural cis-acting elements, and they are often transcript-specific.

- IRE: One of the most well defined structurally based cis-acting elements 3'UTRs is the Iron Responsive Element (IRE) which plays a role in mRNA decay and translation rates [Matoulkova et. al. 2012, Hollams et. al. 2002]. This element is in a 600nt region that contains several different IRE motifs which are usually palindromic in structure [Matoulkova et. al. 2012]. The multiple motifs within the 3'UTR are separated by AREs, which would suggest that this element might play a role in destabilization. In addition, it was observed that within the IRE structure there is a nuclease cleavage motif of "GAAC", further supporting the idea that these IRE would mediate mRNA decay. However, it has been shown that these 3'UTR IREs actually protect transcripts from deadenylation and thus help enhance mRNA longevity by recruiting specific IRE-binding proteins [Erlitzki, Long, and Theil 2002]. This shows a dichotomy of cis-acting regulation vs trans-acting regulation, where a cis-element normally associated with decay can become a stabilizing element due to the interactions of a trans-acting factor.

- SECIS: Another well defined cis-acting element that has a function reliant on its structure is the Selenocysteine insertion sequence element, SECIS, which is a element that allows for the incorporation of selenocysteine into a protein structure [Korotkov et. al. 2002]. Selenocystein is an amino acid that is encoded for by UGA, the stop codon, under specific conditions. In eukaryotes, the structure of the SECIS allows for the binding of SECIS-binding

protein 2 and mediates the incorporation of the selenocysteine. The overall structure of the SECIS was shown to be a large stem-loop that has two internal loops in a small 3 nucleotide loop [Korotkov el. Al. 2002, Mix, Lobanov, and Gladyshev 2007]. Interestingly, amongst eukaryotes the sequence for the SECIS has low conservation, however the structure is highly conserved and has been well established.

2.4 Trans-Acting Factors

Many of the known interactions between proteins and RNA rely on specific protein domains interacting with highly conserved nucleotide sequences. While direct RNA-protein interactions are a common form of trans-acting regulation, other forms require an additional protein that interacts with the initial RNA binding protein to induce an effect [Matoulkova et. al. 2012, Hentze et. al. 2018]. In this scenario the initial RNA binding protein acts as a scaffold for additional elements, but has no direct regulatory function. As our knowledge of RNA-protein interactions has expanded and novel techniques have been developed, it has become clear that RNA-protein interaction do not solely rely on canonical binding domains, but also RNA structure. Most typically, RNA binding proteins (RBPs) can dock on stem-loops. Low-specificity RNA binding in fact bypasses the requirement for sequence specific interactions, such as proteins that bind to the 5'cap. These proteins can bind to any transcript that contains this specific 5' cap element while not needing to interact with a specific motif, therefore widely broadening the scope of mRNA that they can interact with. Furthermore, some specific protein domains are now emerging as important to bind RNA structures, independently of the nucleotide sequence.

The RNA binding protein fragile X mental retardation protein, FMRP for example, contains a RGG/RG motif, that has been identified to be common disordered proteins. This motif interacts with G4-quadruplexes, an RNA structure in which four guanine form a ring like loop where each guanine is engaged with two others through Hoogsteen interactions to form a highly stable, co-planer element [Spiegel, Adhikari, and Balasubramanian 2019]. This interaction

involves the disordered region forming stacking interactions between its arginine and the G4-quadruplex, the flexibility of the glycine then allows the disordered motif to fold along with the G4-quadruplex so the remaining arginine can interact with the G4-quadruplex as well, leading to a strong interaction with high affinity despite no specific sequence, either on the protein or the RNA side [Hentze et. al. 2018].

These sections have defined the importance of regulatory elements in the control of mRNA fate. The complexity and the variability of these elements highlight that regulation of mRNA is a critical aspect of cellular life, where the cell evolved to exploit multiple mechanisms to ensure that a transcript can be properly expressed at specific times and suppressed at others. Perhaps unsurprisingly, viral invaders have evolved in parallel to extensively hijack these mechanics to undermine host processes and decrease gene expression of their host. Furthermore, we now know that viral RNAs can also mimic several of these regulatory pathways, encouraging the host cell to increase expression of viral genes and promote the production of viral progeny.

2.5 Host Shutoff by Gammaherpesviruses: Linking RNA Biology with Viral Hijacking

It is of the utmost importance for viruses to usurp and take control of host gene expression machinery. The evolutionary arms race between viruses and their hosts has led to a myriad of detection system implemented by the host that the virus must overcome. For my thesis, I studied the mechanism primarily used by the gammaherpesviruses endonuclease and their role in inducing widespread mRNA decay. Gammaherpesviruses encode viral endonucleases that belong to the Alkaline Exonuclease family that have both DNAse and RNase function. SOX, the KSHV nuclease, has a highly conserved catalytic domain adopting the classical restriction-like endonuclease motif PD(D/E)XK that can be found in nucleases from bacterial DNA recombination enzymes all the way to the lambda phages nucleases [Covarrubias et. al. 2009,

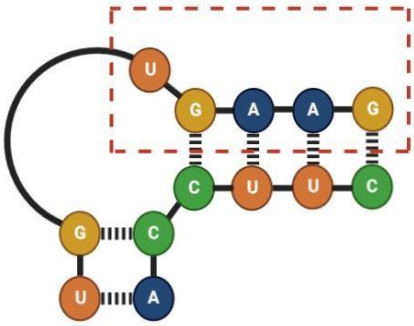

Figure 6: SOX Targeting Motif: Representation of the targeting motif used by SOX. SOX targets the UGAAG sequence when present on the external side of a loop, with cleavage usually occurring at an adenosine within the loop structure.

Buisson et. al. 2009, Kosinski, Feder, and Bujnicki 2005, Covarrubias et. al. 2011].It was shown that these viral endonucleases preferentially target RNAPII transcribed RNA, actively cleave these transcripts, and after the initial viral-induced cleavage event, the host primary exonuclease XRN1 will clear the leftover RNA fragments resulting from the endonuclease cut [Gaglia et. al. 2012]. Early work into SOX highlighted that its expression is essential for viral replication and viral spread [Gaglia, Rycroft, and Glaunsinger 2015, Abernathy et. al. 2015] It was observed that SOX not only targeted host mRNA, but can also cleave any transcript, including viral ones, that contain the SOX targeting motif [Abernathy et. al. 2014, Gaglia, Rycroft, and Glaunsinger 2015]. Despite this seemingly indiscriminate targeting of transcripts, SOX was shown to cleave RNA at specific sites in a sequence-specific manner [Gaglia et. al. 2012, Covarrubias et. al. 2011, Gaglia, Rycroft, and Glaunsinger 2015, Mendez et. al. 2018]. Through transcriptome-wide analysis, it was discovered that SOX preferentially targets a conserved motif common in both human and viral transcripts [Gaglia, Rycroft, and Glaunsinger 2015]. Further investigation found that the core of this targeting motif is a simple 5 nucleotide UGAAG sequence, surrounded by degenerate sequences. This motif adopts a stem-loop structure and SOX appears to preferentially cleave at an unpaired adenosine found within this stem-loop structure [Lee et. al. 2017, Mendez et. al. 2018][**Figure 6**]. Alterations that remove this loop structure lead to reduced catalytic activity, showing that this secondary structure along with the sequence are crucial to SOX targeting.

Virus-Host interplay is a highly dynamic interaction where evolution leads to constant flows and shifts in cellular control. While host shutoff driven by SOX seems to be highly pervasive and dominant, there is significant evidence that some mRNAs are strongly resistant to

its action. This may be a sign that the host system is fighting back, implanting elements that can resist viral nucleases into transcripts as a means to preserve the stability of vital anti-viral transcripts. Investigation into transcript resistance led to the discovery of the first viral-specific nuclease resistance element. This nuclease resistant element represents an unprecedented level of the viral-host battle to control expression as it is ingrained into the structure of select transcripts. Known as the SOX Resistant Element, or SRE, this element has been discovered to bestow resistance to a wide-range of viral nucleases through an as of yet, undiscovered mechanism. Exploration and understanding of this element, both in terms of structural motif and importance as well as mechanism of action, is crucial to understanding the arms race between viruses and hosts. This understanding could pave a way to new therapeutics through the use of artificial transcripts that are resistant to host shutoff, allowing the cell to strike back at the virus. Furthermore, this knowledge will lead to better understanding of RNA stability as this is the first known structural element specifically designed to protect host transcripts from destabilization form a foreign entity.

The SRE has been shown to protect transcripts from a wide array of viral nucleases including all of those within the gammaherpesvirus family. To achieve this, the SRE may act as a binding platform for proteins which block the endonuclease targeting mechanism. Alternatively, the structure of the SRE could allow for the transcript to fold into a more complex organization that does not allow for interactions with the nuclease. Several transcripts have been demonstrated to carry SREs and most of what we know about this “escape” element stems from work done on IL-6. This IL-6 SRE is now routinely used in the laboratory as the “canonical” SRE, but other transcripts such as GADD45B or C19ORF66 were shown to encode RNA sequences with SRE-like functions. Studying this canonical SRE revealed that this RNA element is predicted to fold into a hairpin stem-loop, whose structure appears to be vital to resistance. Yet, these structures are only predicted based on computer modeling. No truly defined structure exists for these elements.

To better understand these structures and to assist in further exploration of their existence in other transcripts, defining the element with a true structure would be a vital tool.

The main goal of my thesis was to better define how the SRE truly folds *in vivo* and *in vitro* and to extend this work beyond the canonical SRE found in IL-6. I will cover my work using a novel chemical probing structural technique known as SHAPE-MaP to elucidate a definitive structure of the SRE. SHAPE-MaP is a technique that exploits the biochemical properties of bound and unbound ribonucleotide to label unbound nucleotides, such as those that participate in secondary structures. Once these labels have been added the sequences are mapped based on these labels to show what regions and nucleotides are forming secondary structures. Through this work we hope to better understand the essential structures involved in host defenses against viral nucleases. We hope to map out a story in which this small structure acts as a the last line of resistance to viral nuclease for anti-viral host transcripts.

CHAPTER 3

MATERIALS AND METHODS

3.1 Plasmids

For *in vivo* expression:

-pcDNA3.1 GFP: This plasmid was used as the base plasmid for cloning 3'UTRs directly downstream of GFP coding region [Schematic representation of this plasmid and all other used in this study can be find on **Figure 7**]. This plasmid was used as the destination vector for all in-fusion cloning to make all qPCR reporter plasmids as follows:

Primers were designed using the TaKaRa In-fusion cloning online tool [See **Table 1** for complete list of primers]. pcDNA3.1-GFP was linearized by restriction digestion using EcoRV and NotI overnight at 37°C and the desired insert was amplified by PCR using KAPA Polymerase following the manufacturer's instructions. Amplified inserts and digested vector were gel purified on 1% agarose gel and gel eluted using the GeneJet Gel Extraction kit and protocol. The In-fusion recombination was performed for 15 minutes at 50°C according to the manufacturer's recommendation. Half of the in-fusion reactions were transformed into D5 chemically competent cells by heat shock and plated onto LB+amp plates. Individual colonies were picked and proper recombination was verified by colony PCR using Sapphire PCR (Takara Bio). Positive clones were then grown over-night in liquid culture, miniprepped, and sent out for Sanger sequencing (Eurofins). This cloning strategy was used to create all vectors that will be used throughout this thesis:

-pcDNA GFP IL-6 3'UTR: The model reporter plasmid that contains the IL-6 3'UTR downstream of GFP. This acts as the negative control for qPCR experiments as we know that the IL-6 3'UTR is sufficient to block SOX-mediated decay.

-pcDNA GFP IL-6 5'UTR: Contains the IL-6 5'UTR region downstream of GFP. As the 5'UTR does not prevent SOX decay, this reporter serves as a negative control for qPCR experiments as we expect to see extensive degradation of GFP.

-pcDNA GFP C19ORF66-3'UTR: Plasmid containing C19ORF66 3'UTR region and is resistant to SOX decay

-pcDNA C19 3'UTR Front Half and Back Half: These plasmids were designed to split the C19ORF66 3'UT in half, with a 11 nucleotide overlap

-pcDNA GFP C19 3'UTR Segments (1-230, 231-460, 461-690, 691-941): C19ORF66 3'UTR was broken down into 4 pieces based off of predictive modeling to avoid disrupting internal structures of the 3'UTR

-pcDNA GFP C19 3'UTR Truncations (1-690 and 1-460): Two mutants designed to ensure that the SRE did not contain elements found in between the segment mutants above. Though predictive modeling did not show structures between segments, these plasmids exist as a means to show structures were not disrupted

-pcDNA GFP BRDT 3'UTR: 3'UTR region of a new potential escapee, BRDT. Obtained from gene synthesis (IDT) and cloned into our reporter system

-pcDNA GFP RAET1E 3'UTR: 3'UTR region of a new potential escapee, RAET1E. Amplified from genomic DNA and cloned into our reporter system

-pcDNA GFP CHRNB4 3'UTR: 3'UTR region of a new potentail escapee, CHRNB4. Amplified from genomic DNA and cloned into reporter system

For *in vitro* expression:

-pBSSK SRE: This Blue Superscript plasmid contains the IL-6 SRE downstream of a T7 Promoter. This plasmid was used in the SHAPE-MaP experimentation for In Vitro Transcription using NEB HiScribe T7 Quick High Yield Kit. Plasmid contains BamHI and NotI restriction sites

to remove SRE and insert others, which was planned to be done if the C19ORF66 SRE was isolated.

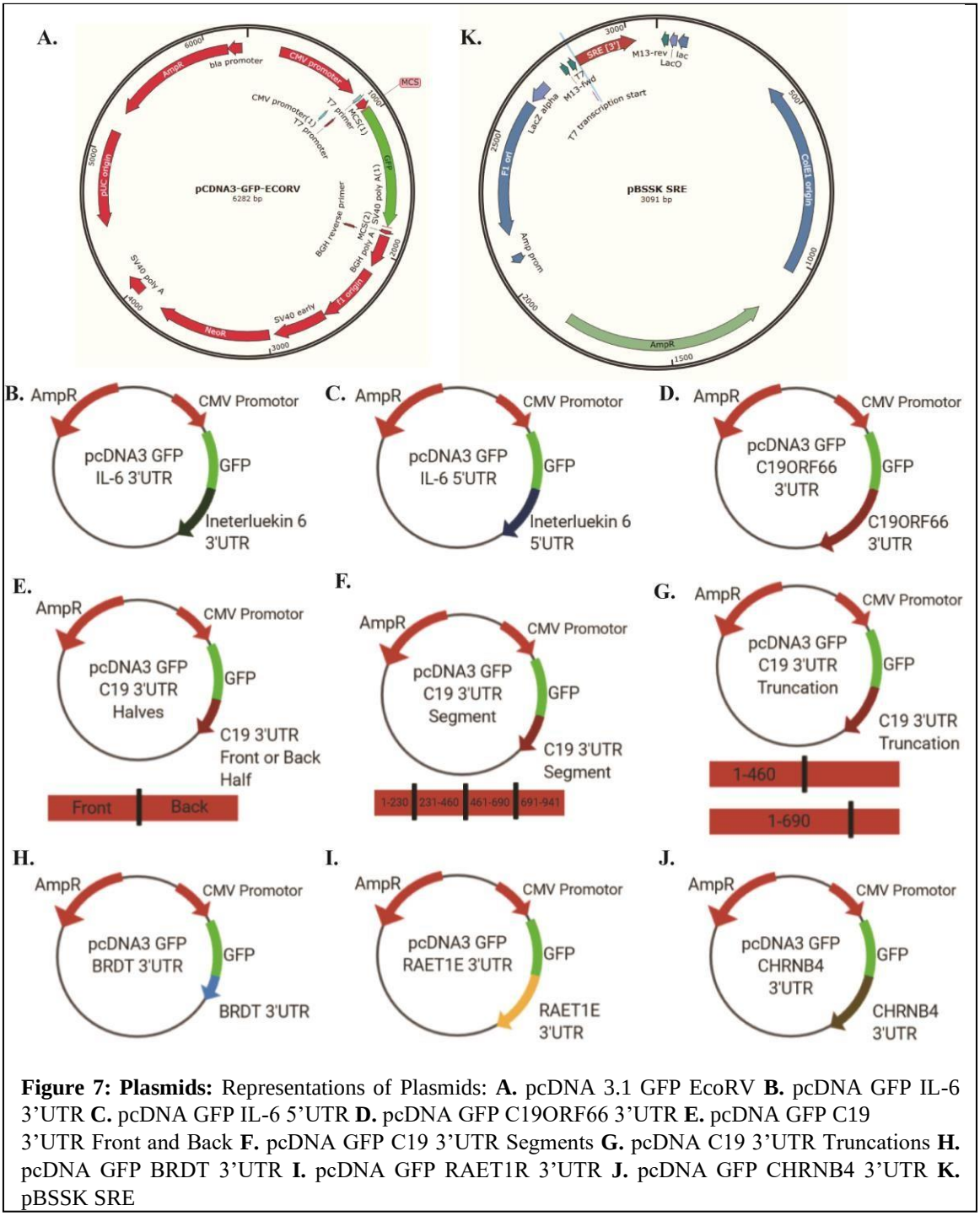

Figure 7: Plasmids: Representations of Plasmids: **A.** pcDNA 3.1 GFP EcoRV **B.** pcDNA GFP IL-6 3'UTR **C.** pcDNA GFP IL-6 5'UTR **D.** pcDNA GFP C19ORF66 3'UTR **E.** pcDNA GFP C19 3'UTR Front and Back **F.** pcDNA GFP C19 3'UTR Segments **G.** pcDNA C19 3'UTR Truncations **H.** pcDNA GFP BRDT 3'UTR **I.** pcDNA GFP RAET1R 3'UTR **J.** pcDNA GFP CHRNB4 3'UTR **K.** pBSSK SRE

3.2 Cells, Transfections, and RT-qPCR

HEK293T cells were grown in Dulbecco's modified Eagle's medium (DMEM - Invitrogen), supplemented with 10% fetal bovine serum (FBS), and 1% Penn/Strep. DNA transfections were performed in 12-well plates: cells were grown for 24 hours to confluency of around 70%. Transfections were performed using PolyJet (SignaGen) following the manufacturer's protocol. 24h post transfection, samples were collected in TRIzol following manufacturer's protocol and either underwent RNA extraction and reverse transcription immediately or were frozen at -80°C for up to a week. RNA extraction was performed via TRIzol chloroform methods. Reverse transcription was performed using AMV reverse transcriptase (Promega) using 1ug of RNA. The cDNA was used for quantitative PCR (qPCR) analysis using SYBR green qPCR (Bio-Rad) on a QuantStudio 3 real-time PCR System.

3.3 SHAPE-MaP

Lab SHAPE-MaP protocol was based on two previously published protocols, Smola et. al. from the Weeks lab and Martin et. al. from the Sztuba-Solinska lab,. *In Vitro* methodology for RNA folding and modification was taken from Smola et. al. and PCR steps and Illumina tagging were adapted from Martin et. al.. This work was done in collaboration with Dr. Joanna Sztuba-Solinska [Auburn University Department of Biological Sciences] who was going to assist with computational and bioinformatic aspects of the ShapeMapper analysis. Sequencing was performed by Dr. Ravi Ranjan of the UMass Genomics Lab. Several iterations of this protocol were developed and the protocol was still undergoing minor modifications for efficiency and consistency when the lab had to shutdown.

3.3.1 Buffers and SHAPE Reagent

- 3.3x Folding Buffer: 33mM HEPES (pH 8.0), 333mM NaCl, and 33mM $MgCl_2$
- Denaturing Control Buffer: 500mM HEPES (pH 8.0) and 40mM EDTA

- 5x Map Prebuffer (also referred to as SHAPE-MaP Buffer): 250mM Tris (pH 8.0), 375mM KCl, 50mM DTT, and 2.5mM of each dNTP.

- RNA Storage Buffer: 10mM Tris (pH 7) and 0.1mM EDTA

- 1-methyl-7-nitroisatoic anhydride, 1M7: SHAPE reagent 1M7 is considered to be versatile and good for general use [Busan et. al. 2019]. 1M7 is an experimental molecule and the longevity is not well understood. The two companies that produce suggest different storage conditions and different "lifespans" for the compound. In this work 1M7 was resuspended in DMSO, actively hydrolyses in water, and stored in small aliquots. Aliquots were treated as having limited freeze thaw cycles due to reagent hydrolysis in water.

3.3.2 Primer Design

Three sets of primers are needed for SHAPE-MaP: Universal RT primer for target zone/transcript, PCR1 primers for adding Illumina adapters, PCR2 primers for Indexing. The RT primer is designed for target zone and has no additional quantifications. RT Primer was designed against the SRE region in the pBSSK SRE plasmid [**Table A1**] PCR1 Reaction adds Illumina adaptors. Base design for the primers is [Illumina adapter]-[5 random nucleotides added during synthesis]-[RT Primer] PCR1 Primers are universal to all experimental conditions. PCR2 Primers are designed to index individual experimental conditions. Specific indexes are used for the different conditions and are underlined in the primer sequence in **Table A1** and are an element of the forward primer. The reverse primer for PCR2 is universal to all reactions and adds the Illumina Universal Adapter which is trimmed off during analysis.

3.3.3 *In Vitro* Transcription

The pBSSK SRE plasmid was linearized using XhoI (NEB) for at least 5 hours at 37°C. After digestion the product was gel purified and extracted. *In Vitro* Transcription was performed using NEB HiScribe T7 Quick High Yield kit following products protocol. Each reaction was: 10ul NTP Buffer Mix, 1ug of Template, 2ul of T7 RNA polymerase mix brought to 20ul with

Nuclease-Free water. After mixing by pulse-spinning, reaction was incubated for 2 hours at 37°C. 2ul of the product was then run on a 1.1% agarose gel to check for RNA Banding. Following *In Vitro* Transcription, samples underwent RNA Cleanup using the NEB RNA Cleanup Kit following company protocol. Samples are raised to 50ul using DNase free water. 100ul RNA Cleanup Binding Buffer is added along with 1 volume (150ul) ethanol for small RNA stability. Samples were eluted as purified RNA in 20ul DNase free water. Purified samples were analyzed by Nanodrop to check RNA concentration to confirm non-zero quantity.

3.3.4 SHAPE-MaP RNA Modification and PCR

To ensure proper RNA folding of secondary structure, samples must first be denatured. 500ng of purified RNA are added to 12ul sterile water. Samples were incubated at 95°C for 2 minutes to denature and rested on ice for 2 minutes. Samples were then supplemented with 6ul 3.3x Folding Buffer and mixed by pipetting and allowed to undergo structural folding by incubating at 37°C for 20 minutes.

RNA modification is done as three different conditions: (+) Modified, (-) Unmodified, and (DC) Denature Control. For the (+) reaction, 1ul of 1M7 was mixed with 9ul of folded RNA and incubated at 37°C for exactly 75 seconds. For the (-) reaction, 1ul DMSO was used instead of 1M7 following previous step. For the DC reaction, 3ul folded RNA was mixed with 5ul of formamide and 1ul of 1x DC Buffer and incubated at 95°C for 1 minute to stimulate denaturing. 9ul of this denatured RNA was mixed with 1ul of 1M7 and incubated at 95°C for 1 minute All reactions were purified using the RNA Cleanup Kit (NEB).

Reverse transcription was performed using AMV RT (Promega) following a modified protocol. For each reaction condition 10ul conditional RNA (+,-, or DC) was used with 2ul RT primer. Primer annealing Reactions were ran under the following conditions: 85°C for 1 minute, 65°C for 5 minutes, and hold at 4°C. RT reactions were setup as follows: a master mix was prepared [4ul SHAPE-MaP buffer, 1ul 120mM $MnCl_2$, 1uk water and 1ul AMV RT per sample]

and 8ul of this Master mix was added to each sample and then incubated at 42°C for 3 hours. NaOH was added to each sample and incubated at 95°C for 5 minutes to hydrolyze RNA template. After incubation, samples were cooled to 4°C and HCl was added to neutralize reaction. Finally, 28ul RNA Storage Buffer is added. Samples were then purified using PCR Purification Kit using spin columns soaked in 100ul RNA storage buffer for 5 minutes. qPCR checks were performed to monitor for sample amplification.

PCR1 was conducted to label sequences with Illumina Adapters. Primers were prepared to 50uM working solution. A master mix was prepared [1.1ul forward primer, 1.1ul reverse primer, 2ul 10mM dNTPs, 20ul KAPA Buffer, and 1ul KAPA Hot Start Polymerase per sample] and 25.2ul of master mix was added to each sample. Final volume of the reactions were raised to 100ul. For PCR2, only 10 cycles were used. Samples underwent PCR Cleanup following GeneJet kit protocol and were eluted at 50ul. Samples were analyzed by nanodrop, and if the analysis was positive for DNA trails were continued into PCR2.

For PCR2 a master mix was not prepared due to primers being specific to trails. For each sample reactions were set up as: 1.1ul specific forward primer (+,-, or DC), 1.1ul universal reverse primer, 20ul conditional PCR1 product (+,-, or DC), 2ul 10mM dNTP, 20ul KAPA Buffer, 1ul KAPA Polymerase, and 54.8ul DNAse free water. For PCR samples were run following the above conditions, but at 10x cycles. Samples were run on a 1.8% gel with wells large enough for 100ul of product. Samples were purified via gel extraction for library preparation.

After library preparation samples will undergo further quantification done by the Genomics lab during the preparation and quantification steps for sequencing. Dr. Ranjan (UMass Amherst Genomics Lab) will perform the following procedures to prepare and sequence the samples: Qubit dsDNA assay and quantification by Bioanalyzer DNA 7500, followed by NGS Library QC to ensure library quality, and MiSeq sequencing using Nano Kit v2 with a PhiX

Spike-in as a control. Following sequencing the data was sent to our collaborator, Dr. Sztuba-Solinska (Auburn University) for final analysis using the ShapeMapper program.

CHAPTER 4

RESULTS

As described in the above sections, our work focuses on the exploration into the mechanisms of the KSHV endonuclease SOX, and the resistance element found within certain transcripts that escape SOX-induced decay. Our past data, based on RNAseq, identified several mRNA that seem to be resistant to multiple viral endonucleases and were also observed to be upregulated during KSHV lytic reactivation [Rodriguez et. al. 2019]. Amongst these transcripts was C19ORF66, an interferon-stimulated gene, also referred to as C19. Early work in the lab went into characterizing the effects that the C19ORF66 protein had on viral-host interplay. Early results showed that since C19 is spared from degradation, the expression of the C19 protein increases over time during KSHV infection. Furthermore, the lab demonstrated that this protein has potent anti-viral properties and stringently restricts KSHV infection. All together, these results highlight the important impact of C19ORF66 on the regulation of KSHV infection and render C19 an interesting potential target for the development of novel anti-viral therapy. However, before further pursuing the role of C19 during infection, it is important to understand the mechanism underlying how C19 escapes SOX-induced decay. Since we know that the C19 3'UTR is responsible for mediating this escape phenotype, I have focused my work on isolating and characterizing the element within this region that provides endonuclease resistance. Early work was also tied to expand beyond C19 into the top 10%

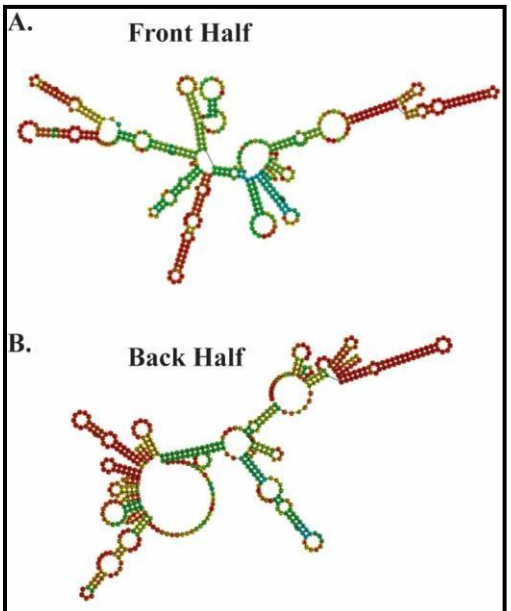

Figure 8: C19 Split Models: Predictive models of the original C19ORF66 3'UTR Split. **A.** Represents the front half of the 3'UTR **B.** Represents the back half of the 3'UTR

SOX-resistant transcripts identified by the RNAseq, but due to cloning complications the results section will focus on the C19 work. The other transcripts, and the complications and means of circumventing them will be explored in the discussion section.

Compared to the 3'UTR of Interleukin-6, the best characterized SOX resistant mRNA, the 3'UTR of C19ORF66 is large: 950 nucleotides in length. Thus, to tackle the problem of narrowing down which region of this 3'UTR is responsible for mediating SOX resistance, we first took a fractionation approach. The C19 3'UTR was split into near equal pieces of 480 nucleotides for the front half and 481 nucleotides for the back half. This created an 11 nucleotide overlap between the two segments to ensure that the two fragments were equal size. The process was based on structural predictive modeling and this overlap helped to avoid disrupting observed predicted secondary structures [**Figure 8** of the two predicted models]. These fragments were then inserted into our GFP reporter as described in the Material & Method section.

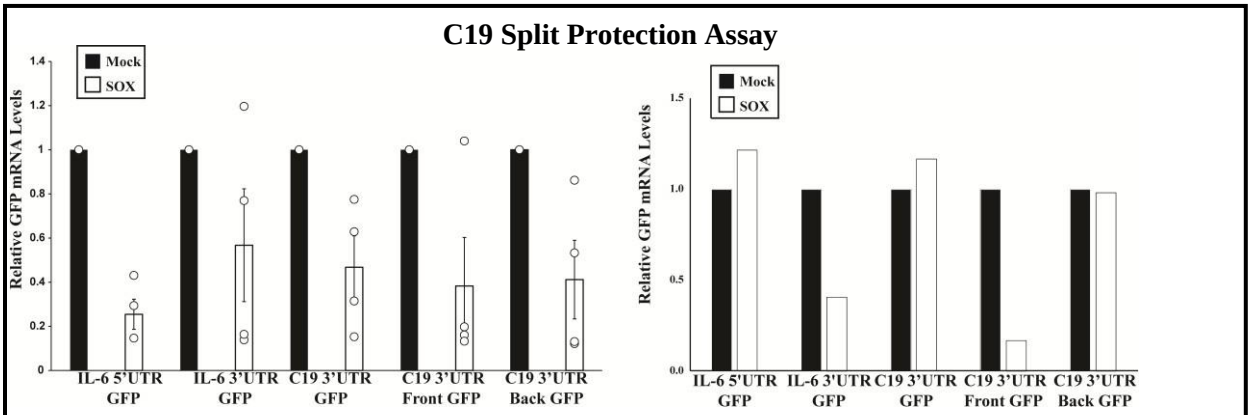

Figure 9: C19 Split Protection Assay: A. Average relative GFP mRNA level of C19ORF66 Front and Back Halves prior to testing and reamplification of endonuclease plasmids. Experiments were erratic and showed degradation of IL-6 3'UTR and C19ORF66 3'UTR in some trials **B.** Final trial of C19ORF66 split after reamplification and testing of endonuclease plasmids. Endonuclease expression plasmids now worked as expected. This test showed GFP protection from SOX by the Back Half of the C19 3'UTR

To test these constructs, 293T cells were transfected with these reporter plasmids along with a SOX (or mock) expression plasmid. To control for SOX activity a positive and negative control are used in all our experiments: a GFP reporter known to be susceptible to SOX (5'GFP) for which we expect to see high levels of degradation, and a reporter known to escape SOX

(GFP-IL-6-3'UTR) for which we expect to see no degradation in the presence of SOX. 24 hours after transfection, total RNA is collected in TRIzol reagent and undergo RNA extraction and RT-qPCR. Data generated from this is analyzed via ΔΔCT methodology. As observed on **Figure 9**, we detect high levels of degradation with our 5'GFP reporter (positive control) upon SOX expression, and no degradation with the GFP-3'UTR (negative control) as expected. The results of these qPCR experiments for the front and back halves of the C19 3'UTR were very variable. Some trials showed low levels of degradation for the front half, other trials resulted in low levels of degradation for the back half, and others still showed high levels of degradation for both halves. This suggested to us that these 3'UTR fragments that we designed were maybe unstable and that our splitting approach was not ideal to find the escape element within C19 3'UTR.

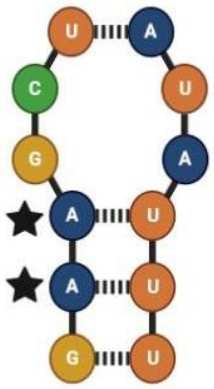

Figure 10: IL-6 Internal Loop Structure: Schematic of the internal loop of the IL-6 SRE with mutation of the marked nucleotides leading to loss of resistance

While testing these constructs, I began working with an undergraduate student, Isiaha Price from Amherst College. Since the C19 3'UTR split experiments were proving inconclusive and the escape element had yet to be identified, together, we began investigating which fragment of the C19 3'UTR was responsible for escape with a greater consideration to secondary structures within the predicted model. Past experiments using the SRE from other escaping transcripts had shown that SOX resistance was linked to a prominent hairpin structure. Mutation of two nucleotides within this hairpin was shown to abolish resistance. These two nucleotides were within the sequence GAAGC, where the GAA was located at the start of an internal bulge and the GC were part of the internal bulge [schematic in **Figure 10**]. The initial experiments disrupted the AA sequence, which was predicted to completely disrupt the structural formation. Our investigations of the C19 3'UTR did not identify this sequence. However, we identified several prominent stem-loop structures that shared structural similarities to the established SREs. We thus decided to split

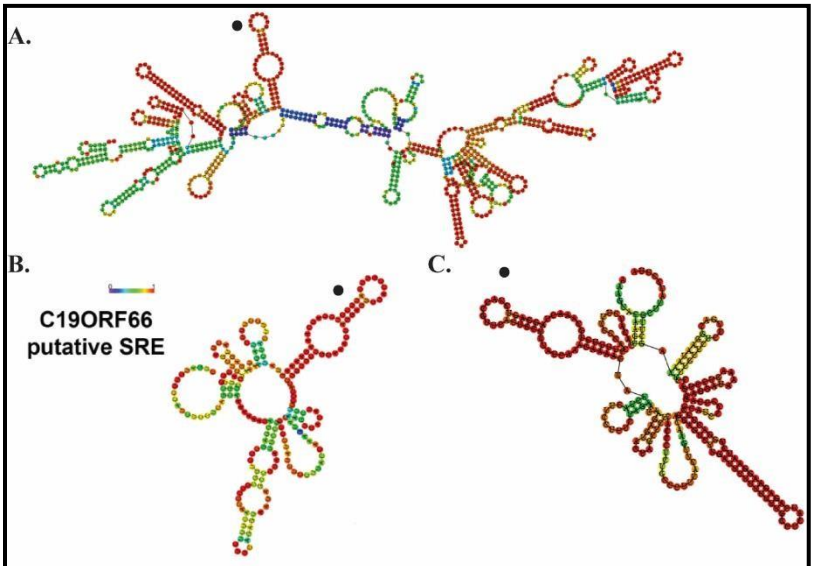

Figure 11: C19 SRE Comparison: Predicted models of: **A.** Full Length C19ORF66 3'UTR **B.** Putative C19ORF66SRE adapted from Rodriguez et. al. based on sequence alignment to IL-6 SRE **C.** C19ORF66 3'UTR Segment 691-941 Each of these structures contains the same stem-loop, suggesting that this structure is highly conserved and a probable SRE target

the 3'UTR into four smaller fragments as nucleotides 1-230, 231-460, 461-690, and 691-941 along with two truncations of 1-690 and 1-461. In particular, the 691-941 segment appears to contain a stem-loop structure reminiscent of previously identified SRE structures and this secondary structure is conserved within the full length 3'UTR predicted model, as well as the full C19ORF66 predicted model [**Figure 11** Comparison of putative structure to 691-941]. All constructs were made and sequenced.

However, due to the laboratory shutdown, due to Coronavirus-19, only one qPCR trial was performed on these constructs prior to lab closure. These results were promising as only one segment was observed to mediate escape from SOX: as shown in figure, fragment X appears to be refractory to SOX cleavage while all the other fragments are susceptible, strongly suggesting that

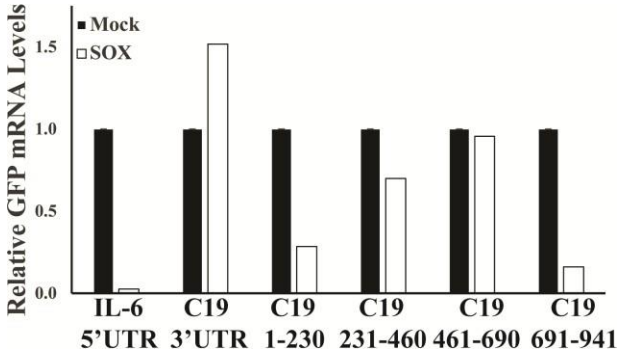

Figure 12: C19 Segment Protection Assay: 293T Cells transfected with a GFP reporter plasmid containing one of the segments of the C19 3'UTR and either a Mock expression or a SOX expression vector. Representative of relative GFP levels. If a segment provides protection, the GFP levels should be near or above 1. IL-6 5'UTR and C19ORF66 3'UTR are controls

fragment X encompasses the putative C19 escape element [**Figure 12**]. The data shows minimal degradation of the 461-690 fragment, suggesting that this region is providing resistance to SOX. Future work on these constructs will be finished by Isiaha for his undergraduate thesis.

While working to narrow down the resistance region important to the escape mechanism of C19 was a major part of my work, the core of my research was to reveal not just the sequence, but the mechanism behind escape. As RNA structures play a critical role in mRNA turnover, both directly and indirectly, we began to investigate the RNA escape element to determine its secondary structure. We know that the structure of the element is crucial to its function as an endonuclease resistance element and working based solely on a predictive model is not enough. For this reason we began to work in collaboration with Dr. Joanna Sztuba-Solinska to use a chemical probing structural technique SHAPE-MaP to characterize the secondary structure of the SRE. SHAPE-MaP is a powerful technique for RNA structural work as it combines sequencing methodology with chemical probing of unbound nucleotides. This overcomes many of the difficulties when it comes to determining RNA structures. While we plan on eventually generating *in vivo* structures, or initial work was focused on performing *in vitro* experiments as a proof of concept.

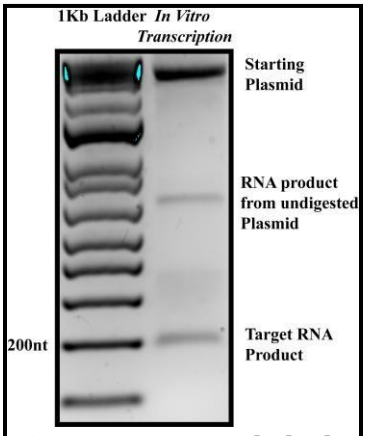

Figure 13 In Vitro Gel Check: Gel Electrophoresis of *In vitro* transcription product. The right lane shows three bands: the top band is indicative of template, the middle band is likely an undesired RNA product from undigested template, and the bottom band is the target RNA product with the expected product size being ~202 base pairs.

To look at the SRE *in vitro*, we started by optimizing our *in-vitro* protocol. After plasmid linearization, products were gel purified and underwent *in vitro* transcription using a T7 polymerase based kit. Initial *in vitro* transcription products were run on gel to ensure production of product [**Figure 13**]. As shown in **Figure 13**, three bands are visible, the top band is the template, the middle band is likely an RNA product that is due to

the presence of circular starting plasmid being present, and the lowest band is the target SRE RNA at the expected size. This confirmed that *in vitro* transcription was working. Following *in vitro* transcription, we started optimizing the RNA folding and modification steps of this protocol.

RNA folding was done using a buffer that mimics biological folding conditions by providing the proper ions and pH. RNA modification was performed using the SHAPE reagent 1-methyl-7-nitroisatoic anhydride, 1M7, which is highly reactive to the presence of water. Modification steps need to be performed rapidly with high accuracy to prevent accidental early hydrolysis of the reagent. After modification samples underwent reverse transcription following the conditions set in the material and methods. These conditions allow for the polymerase to substitute in random nucleotides at modified nucleotides, which is used to determine structural interactions. To test that samples had been properly converted to cDNA, a fraction of the samples was used to assess RT efficiency by qPCR. As shown in figure X, we detected three distinct amplifications from each sample, confirming the production of cDNA from RT-PCR [**Figure 14**]. The differences in cycle threshold (ct) reflects how much RNA was present in the samples: because the modified and denatured samples have lower concentrations of RNA, we expect them to take longer to amplify and thus have higher ct values as observed here. This style of amplification plot was observed for each trail where a qPCR check was performed with similar distances between trials.

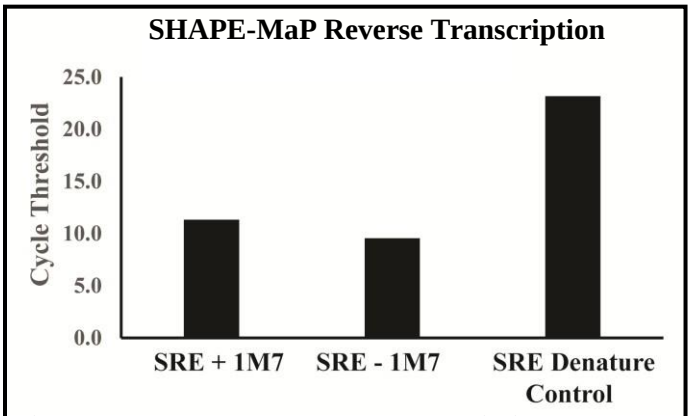

Figure 14: SHAPE-MaP Reverse Transcription Test: qPCR check of SHAPE -Map Samples to check for number of cycles necessary to be detected for confirmation of successful reverse transcription

After reverse transcription, samples underwent two rounds of PCR amplification. The first round, PCR1, is designed to add Illumina adapters to the newly made cDNA in all of the

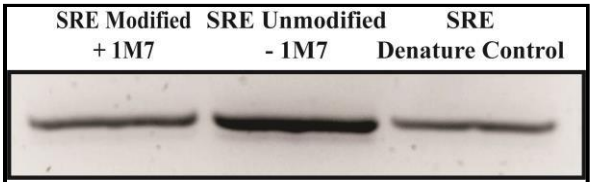

Figure 15: SHAPE-MaP Library Preparation: Final Library Preparation of samples from SHAPE-MaP trial. The left lane is the 1M7 Modified SRE, the middle lane is the Unmodified SRE, and the right lane is the Denature Control SRE.

reaction conditions. PCR2 supplements this by attaching Illumina Indexes for sequencing. This process is limited to 30 cycles between the two steps to avoid amplification bias. Amplification bias involves the development of extra copies existing in one read when compared to another read. Structural generation in SHAPE-MaP requires the data produced from all three conditions to create an accurate structure. If one structure is represented more heavily, due to higher amplification rates or more mutations due to increase cycles compared to other conditions, then this bias might alter if a nucleotide is marked as paired or unpaired in the final structure. After both rounds of PCR the samples undergo library preparation by gel electrophoresis and gel extraction. **Figure 15** shows a completed library preparation with the expected results of all three conditional bands. This is indicative of full PCR amplification of the SRE target under each of the experimental conditions, if there were complications at any step that prevented amplification of the target then no band would be visible.

Sequencing was performed by the Genomics Resource Lab operated by Dr. Ravi Ranjan. This was performed through Illumina sequencing using the MiSeq Reagent Nano Kit v2 a 500 cycle kit as the primary methodology. Before sequencing the Genomics Lab provided quality control through a Qubit dsDNA assay to test sample concentration supported with a DNA high sensitivity assay through a Bioanalyzer if the concentrations were deemed too low. This was followed by a NGS Library quality check in the form of a qPCR assy. The sequencing was controlled for with a PhiX Spike-in for a low diversity library, as there are only three conditions to each trial. Once the Illumina BaseSpace data was received the data was shared with the Sztuba-Solinska lab for analysis via ShapeMapper. This program used all three conditions to provide a detailed model by looking at the rates of variation due to mutation. The modified condition will

have mutations based both on SHAPE reagent modification, as well as normal mutation rates due to polymerase activity. The unmodified condition modulates for SHAPE reagent induced mutation rates and acts as a comparative control. While the denatured control is used to correct for intrinsic background mutation rates from reverse transcription. This data allows for accurate structuring of secondary structures by mapping out the SHAPE modified regions as unbound regions. From this a accurate model will be produced that maps out the desired structure with high accuracy.

Results of the ShapeMapper program generated several different data sets which were used in totality to accurately develop the final SHAPE model. **Figure 16A** shows a map of possible pairings for each nucleotide based on the windowed folding analysis used by the program. For the SRE this shows that there is a high degree of nucleotide interactivity within the structures, which suggests a high degree of flexibility throughout the structure. Further analysis, including measurements of Shannon Entropy compared to reactivity was used to determine base pairing partners represented in **Figure 16B** which shows the base pairing interactions within the IL-6 SRE. From these measurements and the chemical probing interactions a quantitative structure was generated as shown in **Figure 16C.**

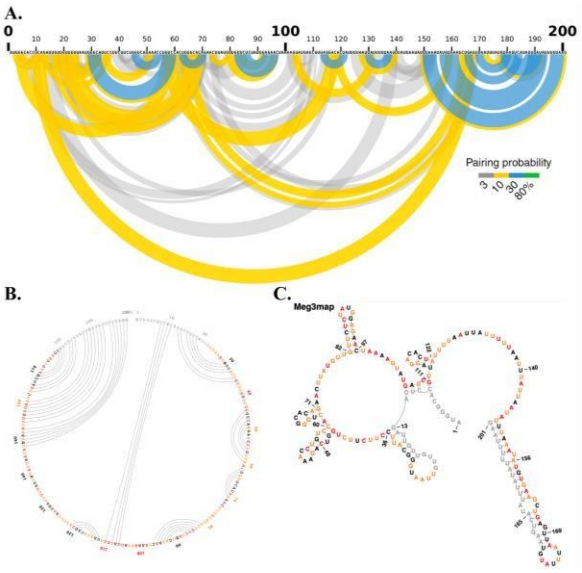

Figure 16: SHAPE Data: Data generated from the SHAPE-MaP experimentation and ShapeMapper analysis. **A.** Map displaying SRE region nucleotide sequence and base pairing probability. **B.** Circular representation of SHAPE model based on reagent reactivity and Shannon Entropy, with lines showing binding partners, different colored nucleotides represent different SHAPE reactivities. **C.** SHAPE model of IL-6 SRE

CHAPTER 5

Discussion

Host shutoff, as mediated by viral endonucleases, is a significant turning point in the interplay between viruses and their hosts. Through the widespread destruction of host transcripts the virus mediates host immune response and usurps host machinery for the production of viral transcripts, replication of viral genomes, and the construction of progeny [Jenner et. al. 2001, Gwack et. al. 2001, Zhu and Yuan 2003]. However, despite the widespread destruction of host cellular mRNA, a small percentage of host transcripts are able to escape this event [Glaunsinger and Ganem 2004, Clyde and Glaunsinger 2011]. Within a subset of these escaping transcripts, recent studies have isolated a resistance element that actively protects mRNA from viral endonuclease induced decay [Hutin, Lee, and Glaunsinger 2013]. This resistance element (referred to as the SOX Resistant Element or SRE) was demonstrated to be contained within the 3'UTR and identified as a prominent hairpin loop secondary structure within the Interleukin-6 transcript [Muller and Glaunsinger 2017]. While the Interleukin-6 SRE acts as our model for the existence of viral endonuclease resistant RNA structures, only one other well defined structure has been heavily documented in the GADD45B transcript [Muller and Glaunsinger 2017]. Previously we had identified 75 host transcripts, through the use of RNASeq, that escape viral-induced endonucleolytic cleavage that may contain RNA escape elements [Rodriguez et. al. 2019]. In this work we further explored one of the most prominent of these escapees, C19ORF66. Beyond this, we worked to advance our understanding of the IL-6 SRE through chemical probing techniques to develop a true experimental model of the resistance element. Based on our understanding of the currently identified SREs, we hypothesize that within the 3'UTR of the C19ORF66 transcript that there will be a prominent hairpin loop structure that is essential to endonuclease resistance. Furthermore, we expect to garner further understanding of how the SRE protects the host transcript through our experimental model. Through this work we will advance

our understand of the interplay between RNA structure and viral protein interactions by furthering our understanding of the structural significance of the SRE.

As shown previously in **Figure 9**, while the back half of the C19 3'UTR seems to provide some levels of protection from SOX, it is not enough to fully protect our GFP reporter. Several reasons could explain this inconclusive result: first, as shown on the bar graph, we had extensive heterogeneity in our results. We tested several plasmids (both for our 3'UTR constructs and the endonucleases), several types of tags, fresh cells and yet, continued to observe the same level of heterogenicity. We are still troubleshooting this aspect of my project, for those who will follow in my work. Secondly, one possibility is that by splitting the 3'UTR in the middle, as we did, led to the disruption of secondary structure and thus directly affected the stability of our construct, independently of SOX activity. Therefore we are hopeful that our more recent effort to design fragments based on predicted structure will yield more decisive results. In particular, our preliminary results indicate that at least one fragment only sees minor degradation in the presence of SOX suggesting that we may be able to fully isolate the resistance element in C19. As described earlier, this work will be carried out by Isiaha for his senior thesis since our work was cut short by SARS-CoV2.

While the work on the C19ORF66 3'UTR and the structural work on the IL-6 SRE made up the bulk of my project, I was also interested in exploring other newly identified escaping mRNA. For this I worked on attempting to amplify the 3'UTRs of several transcripts including: BRDT, RAET1E, CHRNB4, and ARMC10 to name a handful. These transcripts were selected based on their ranking by RNAseq (keeping only high ranking/high confidence escapees) and their known function (focusing on transcripts with putative functions in viral-host interplays). However, from the start, we experience a lot of technical difficulty in this project: cDNA heterogeneity in cells made designing primers challenging, low transcript expression made it hard to isolate RNA to achieve any level of PCR amplification from cDNA, high GC content as well as potential complex secondary structures in these UTRs rendered most polymerases ineffective.

All of these problems also mostly hindered our ability to order these UTRs as synthetic gene fragments, as high GC content interferes with synthesis. To address these problems, I continued to look for a method through which we could successfully clone the 3'UTRs of these transcripts. Early attempts included using additives, such as betaine and ethylene glycol, to the PCR reactions. These additives have been shown to assist in overcoming translational stalling by disrupting GC interactions and weakening secondary structures. Moreover, I also attempted to use temperature gradients to allow for better binding of primer and to assist in the breaking down of secondary structures found within the 3'UTRs. However, none of these attempts yielded any satisfactory results. Finally, we decided to try to use genomic DNA as the template instead of cDNA. Using genomic DNA is obviously not ideal to clone UTRs: many RNA processing events determine which portions of the genomic DNA is turned into cDNA so we might bias our product. However, by catching these sequences prior to RNA folding, we might increase our chances for primer binding and amplification. The logic behind this methodology is that the 3'UTR is coded within the last exon, though the last exon may contain additional sequence information. At the same time, the genomic DNA does not contain many modification related to secondary structure as the sequence need to be tightly packaged and rapidly unwound. I decided to try this technique on RAET1E as its 3'UTR was fully in the last exon with only 10 or so additional nucleotides. I designed primers for the last exon and performed PCR using HEK-293T cell gDNA as a template. Excitingly, the sequence was fully amplified on the first trial, and after infusion cloning, was inserted into our GFP reporter. Further sequencing proved that the sequence was correct and we have thus successfully cloned this UTR. Based on this result, I continued with the CHRNB4 3'UTR using primers targeted to the 3'UTR and not the last exon. Once again, this worked successfully. The primary problem surrounding the study of potential escapees had been overcome. Before the lab shutdown due to Coronavirus, I was working with my undergraduate to teach him to extract gDNA and amplify the 3'UTRs of escapees through the use of gDNA as his undergraduate thesis will focus on the isolating the SRE of several of our candidates. This opens

a huge door of potential for research in the lab as we can begin to generate a library of 3'UTR containing reporters to test the escape potential of our targets.

SHAPE-MaP is an RNA based technique that required learning a significant amount of information on RNA biochemical properties and interactions. The process takes advantage of the unique principles that define RNA in paired and unpaired interactions within RNA structure. SHAPE stands for selective 2'-hydroxyl acylation analyzed by primer extension with the MaP portion standing for mutational profiling. This is due to the focus of the methodology focusing on the reactivity and flexibility of the 2'-hydroxyl group found within RNA [**Figure 17A]**. When an RNA molecule participates in a binding interaction with another nucleotide the 2'OH group is placed in an orientation where it becomes ridged and not prone to participate in reactions. However, when a RNA nucleotide is unbound, the 2'OH group can enter a transitional state where it is acted upon by the 3'phosphodiester within the backbone. This interaction is catalytic and allows for interactions between the OH and a strong electrophile. SHAPE-MaP exploits this. SHAPE-MaP experiments use a reagent that interacts with flexible hydroxyl groups as a strong electrophile. This means that the reagent is able to come in and participate with an element that is normally not highly reactive because of the flexibility and catalytic state. In doing so, the reagent modifies the hydroxyl group into a large, bulk adduct on the 2' location. In our work we used 1-methyl-7-nitroisatoic anhydride [**Figure 17B]** which is considered to be the most versatile of the SHAPE reagents and is primarily used cell-free conditions with small diversity libraries. This means that we are able to readily modify our RNA targets *in vitro* without complications due to our reagents and can expect the reagent to act equally amongst all transcripts within our samples. Beyond this 1M7's low half-life of 17 seconds, and its self-limiting actions, meaning it hydrolyzes rapidly with water when no thermodynamically favorable partners remain make it an easy reagent to work with. Compared to another *in vitro* SHAPE reagent, N-methylisatoic anhydride which has a 260 second half-life, we do not have to worry about RNA degradation due to prolonged exposure to room temperature or the possible interactions of the reagent with less

thermodynamically beneficial partners. Therefore, we can expect no interactions with RNA molecules that are participating in paired interactions which means no mislabeling.

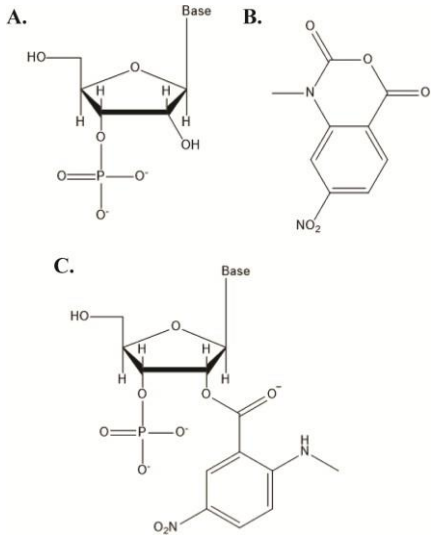

Figure 17: SHAPE Reaction: Representations of elements involved in SHAPE reaction **A.** RNA molecule with flexible 2'OH group **B.** 1-Methyl- 7-nitroisatoic anhydride (1M7) **C.** Modified RNA Molecule with 2'O-adduct

Following labeling with the bulky 2'O-adduct [shown in **Figure 17C**], reverse transcription is used to introduce mutations where 1M7 added the 2'O-adducts. When the reverse transcriptase reaches a nucleotide that has been modified it is unable to read the present nucleotide. However, the modified reaction conditions, the presence of DTT in the buffer along with other additives, allows for the transcriptase to substitute in a random nucleotide causing random mutations in the modified structure. The following PCR steps then amplify these mutations while adding the sequencing adaptors and indexes to create a mutant library which is ultimately sequenced. This process allows us to rapidly identify which nucleotides in an RNA transcript are participating in base-pair interactions and which are not. Analysis by the Shapemapper program compares these modified, mutant libraries to the unmodified library mapping out where mutations occur to chart paired and unpaired nucleotides. These results are compared with the denature control condition to take into account the intrinsic mutation rates of the polymerases. The totality of this data is then used to map out a structural profile based on the known sequence, the locations of base-pair interaction or the lack of interactions, and folding entropies to form an accurate model based on experimental data.

This mutational profiling data along with the computation modeling with error estimates allows for the generation of models for RNA structures. While these models are still based on algorithms for structural modeling, the addition of the chemical probing to identify secondary

structures has yielded a methodology that produces highly accurate models that are comparable to known structures. Many of the early SHAPE and SHAPE-MaP experiments were done using RNAs with known structures to test the accuracy of this modeling. The inclusion of SHAPE data was shown to significantly increase the accuracy of computer modeling to the point tat the generated models were accurate to the structure of these known RNAs [Siegfried et. al. 2014]. These experiments, and the comparative ease of SHAPE and SHAPE-MaP when compared to physical modeling methods, make the process a strong candidate for RNA modeling when exploring secondary structures and interaction.

Many complication arose during my SHAPE-MaP work as I worked to create a lab protocol. The initial trial did not produce a denature control band during library preparations. However, the underlying problem was adjusted for and corrected to generate a full conditional library for the second trial with little difficulty. Work to produce a third trial to have triplicate data led to several complications. After several additional trials, only the unmodified condition was producing banding in our library preparation. We concluded that our SHAPE reagent must have expired as it is an experimental compound with little information on storage protocols and life span. We also concluded that the reagent may have hydrolyzed itself due to the presence of ice during several freeze thaw cycles. We ordered a fresh supply of 1M7 and after resuspension in DMSO separated the reagent into several small aliquots, just in case freeze-thaw was the culprit. The next trial showed faint banding for all of the conditions, confirming our suspicions. As PCR2 does not use all of the PCR1 product, we went back to the PCR1 product and redid PCR2 as a rerun trail and received slightly brighter banding. It was during the preparation of a new replicate that the lab was shutdown due to the Coronavirus Pandemic. Following the partial reopening of the lab, the sequencing has been completed with the assistance of Dr. Ranjan (UMass Amherst Genomics Resource Laboratory Director) and the sequencing data was sent to our collaborator

Dr. Joanna Sztuba-Solinska. The data was processed with the Shapemapper program to generate a model based on the SHAPE-MaP experiment data [**Figure 18**].

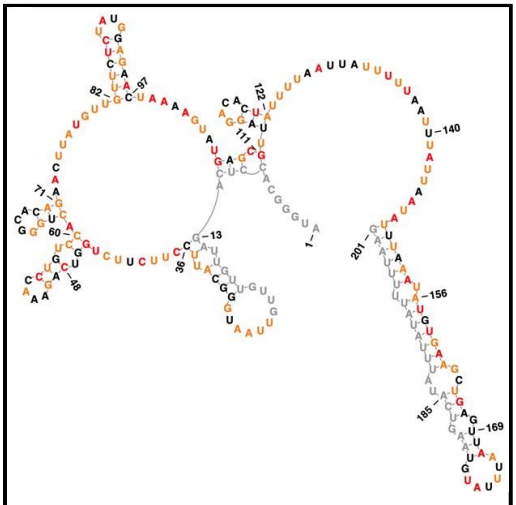

Figure 18: IL-6 SHAPE Model: Model of the IL-6 SRE generated from the SHAPE-MaP experimental data. Color is indicative of SHAPE reagent reactivity where black nucleotides have low reactivity, yellow have slight reactivity, and red are highly reactive to SHAPE reagent

The most striking aspect of the generated model is the presence of the predicted stem-loop that was previously shown to be essential for the endonuclease resistance. Where as previously we expected this region to be some sort of stem-loop essential to the function of the SRE we now have an experimentally supported model that proves this structure exists as we expected. Another aspect we could not have expected from the model is just how open this structure is. From looking at the coloration within **Figure 18** and supported by **Figure 16A**, we can see that the vast majority of nucleotides in the structure are highly reactive to the SHAPE reagent, indicating that these structures are highly reactive and structured. Elements like the stem-loop are highly reactive indicating high structural flexibility, indicating that the structure is likely to constantly be in flux with bonds breaking and reforming. This indicates a high degree of flexibility, that the structure is able to form dynamic interaction by having a significant degree of open nucleotides for interactions. As suggested by *Smola et. al.* this may indicate that this region could be involved in interactions with proteins, which would support the SRE acting as a binding platform for protective protein elements. As these reactive areas are able to open up and form interactions with other cellular elements. This open flexibility could allow the development of a large protein scaffold that protects the SRE from SOX and other endonucleases. With this model, we now know for certain that the SRE is a

highly dynamic site, capable of a wide degree of flexible interactions, supporting our belief that the SRE is responsible for the recruitment of a protein complex as part of the protection element.

While producing a lab protocol for SHAPE-MaP several of our initial decisions were changed. Whether or not to do phenol chloroform extractions of the digested starting plasmid as well as certain checks and balances to make sure steps were working as intended. Several of these were either done only in the beginning before full trials were performed as a proof of concept. Amongst these the qPCR check after RT was one of the most discussed steps. Originally we came to the conclusion that we would see three different amplification plots due to all three conditions amplifying at different rates [as seen above in **Figure 14**]. We predicted the unmodified would replicate the fastest, the modified after that, and then the denatured at a much higher cycle threshold later. This was observed for each qPCR check performed. However, we began to doubt this as an accurate means of testing as differences in concentration, due to modification or the denaturing step, may influence this instead of actually predicting if modification and denaturing had been successful. Furthermore, certain gel checks, such as for RNA production from *In Vitro* Transcription, required such high amounts of the product to produce visible banding it was not a viable check to perform during a SHAPE-MaP trial. As such the early gels after *In Vitro* Transcription act as a proof of concept to show that our methodology is indeed producing an RNA product at the size we expect, but is not a reliable method for confirming RNA production every trial.

As a procedure SHAPE-MaP is difficult and requires fine tuning. However, I have no doubts that the process is simpler and more pertinent to the study of large amounts of transcripts when compared to X-ray crystallography or Cryo-EM which are better suited to the study of tertiary interactions of biologically significant RNAs, such as tRNA and rRNA. If our trials from the IL-6 SRE produce viable structural data, then I can foresee this process being used to explore other SRE structures, especially that of C19. It is a highly valuable technique for studying RNA

structure as the protocol is relatively simple and can be modified for both *In Vitro* and in cell work. The methodology represents a significant advancement for the study of individual RNA, both coding and noncoding, and will likely be used in the discovery of many novel functional RNA structures. However, the transfer from *in vitro* to *in vivo/in cellulo* work poses a drastic change in local environment and with this comes new struggles and potential complications. In a living cell RNA are often engaged in interactions with a wide array of molecules, from the ions that help stabilize cellular environments to the proteins that interact with transcripts. The presence of such elements has been shown to alter SHAPE reactivity, which may lead to different final structures [Martin et. al. 2019]. This would show the largest variation within single stranded regions where the interactions of proteins or ions could lead to a lower SHAPE reactivity which could lead to the process labeling these nucleotides as participating in base pairing. Additionally different SHAPE reagents may be necessary for in cell work. While 1M7 has been used in cells, its low half-life may make it more difficult to accurately and fully modify all transcripts as cells are extremely busy, crowded environments. To overcome this, larger quantities of 1M7 could be used, but this may cause bias. To further complicate this, the target sequence or transcript may exist at low basal levels, which could make modification and isolation of the target transcript more difficult. This can be supplemented by the use of an expression vector, but there exists the possibility that this expression vector might not be identical to the cellular produced RNA in terms of modifications and thus differences may exist between the determined model and the true structure. From performing SHAPE-MaP *in vitro* we hope to garner a better understanding of the process and understand where complications may arise in the process. We have already made adjustments to the means by which we keep and handle 1M7 and have learned how to develop checks to make sure that a procedure is working. When we proceed to *in vivo* we will have a better understanding of the fundamental principles making troubleshooting easer, as we have learned how to identify issues pertaining to mistakes in PCR vs the SHAPE reagent not functioning properly. With this powerful tool we hope to unravel some of the intrigue

surrounding the function of SRE. Is there some internal element of the SRE that predictive modeling does not form? Maybe we will have data better suited to the identification of protein binding sites that can be used in conjunction with mass spectrometry data that is being generated from another branch of the lab to determine which proteins are directly interacting with the SRE.

RNA is a demanding molecule to work with as it is highly sensitive to environmental conditions and readily degraded by simple interactions. Despite the complications and difficulties working with this molecule advances in RNA biology are elucidating the significance of many aspect of RNA that were once taken for granted. One of these aspects is the importance of secondary structure in RNA function and fate. In this thesis we explored how a novel set of structures, the SOX Resistance Elements, provide increased stability and resistance to viral nucleases. In addition, we made significant advances in mapping out the exact structure of the model example isolated from the IL-6 3'UTR. Additionally we laid the foundation for future work identifying and isolating additional resistance elements that may be concealed within our potential escapees. From this work we will garner a better understanding in the interactions that allow viruses to hijack host systems and devise therapeutic methodologies for manipulating these interactions by using elements such as the SRE to disrupt viral shutoff. These elements could be used to develop artificial transcripts that are resistant to shutoff while being able to trigger cellular immune responses allowing for virally infected cells to undergo apoptosis. This potential is reliant on our understanding of the mechanisms involved within these pathways, and by generating a true structural map for our SREs we may discover some unknown facet that could hold the answer to how these elements overcome viral endonucleases.

APPENDIX A

TABLE A1: PRIMER LIST

Target	Forward Primer	Reverse Primer
C19 3'UTR Front Half	TCAATGTGTAGCGATCCCTGCCAGGTGCAGATACAAACC	TACTCTAGAGCGGCCAACTGTGCCCCTGTGCCC
C19 3'UTR Back Half	TCAATGTGTAGCGATG1GGGCACAGTTACCTGCAG	TACTCTAGAGCGGCCTATAAGGATAAGCAAGCTTTTATTG
C19 3'UTR 1-250	TCAATGTGTAGCGATCCCCTGCCAGGTGCAGAT	TACTCTAGAGCGGCCTTGTTCTGGCCACACCTTCT
C19 3'UTR 251-460	TCAATGTGTAGCGATCTTGGGCTCCTGCTGACC	TACTCTAGAGCGGCCCAGCGGGGAGGCTGGTGG
C19 3'UTR 461-690	TCAATGTGTAGCGATGTACAGGGCACAGTTACCT	TACTCTAGAGCGGCCGGTAAGGAGATAGGGAAGGAAT
C19 3'UTR 691-941	TCAATGTGTAGCGATAAAGTACAAGTCACATCTTTCCC	TACTCTAGAGCGGCCTAAGGATAAGCAAGCTTTTATTCCG
C19 3'UTR 1-690	TCAATGTGTAGCGATCCCCTGCCAGGTGCAGAT	TACTCTAGAGCGGCCGGTAAGGAGATAGGGAAGGAAT
C19 3'UTR 1-460	TCAATGTGTAGCGATCCCCTGCCAGGTGCAGAT	TACTCTAGAGCGGCCCAGCGGGGAGGCTGGTGG
SHAPE-MaP RT	CTTAAAAATATAAATATGACTTACATAAATTAAC	
SHAPE-MaP PCR1 WT	GAC TGG AGT TCA GAC GTG TGC TCT TCC GAT CT NNNNN ATGGGCACCTCAGATTG TT	CCC TAC ACG ACG CTC TTC CGA TCT NNNNN CTTAAAAATATAAATATGAC
SHAPE-MaP PCR2 +	CAA GCA GAA GAC GGC ATA CGA GAT TGACCA GTG ACT GGA GTT CAG AC	AAT GAT ACG GCG ACC ACC GAG ATC TAC ACT CTT TCC CTA CAC GAC GCT CTT CCG
SHAPE-MaP PCR2 -	CAA GCA GAA GAC GGC ATA CGA GAT GCCAAT GTG ACT GGA GTT CAG AC	AAT GAT ACG GCG ACC ACC GAG ATC TAC ACT CTT TCC CTA CAC GAC GCT CTT CCG
SHAPE-MaP PCR2 DC	CAA GCA GAA GAC GGC ATA CGA GAT CTTGTA GTG ACT GGA GTT CAG AC	AAT GAT ACG GCG ACC ACC GAG ATC TAC ACT CTT TCC CTA CAC GAC GCT CTT CCG
BRDT 3'UTR	TCAATGTGTAGCGATTCAATGTGTAGCGATAACTCAG	TACTCTAGAGCGGCCTACTCTAGAGCGGCCTGTAAG
RAET1E 3'UTR	TCAATGTGTAGCGATGCAGAAGATCCACCTAGAGG	TACTCTAGAGCGGCCTGCATTCAGTGTTAAAGTGTTTATC
CHRNB4 3'UTR	TCAATGTGTAGCGATGGGCCCCCTGGGTTGTGG	TACTCTAGAGCGGCCACAAACATTTATTGAGCACCTACTG

APPENDIX B

3'UTR CLONING TRIALS

From the start, the 3'UTR cloning experiments provide many technical difficulties. Due to the heterogeneity in the cells primer design was challenging and this along with low transcript expression made it difficult to isolate RNA targets for our potential escapees. Furthermore, high GC content, along with the potential complex secondary structures that we were searching for within these regions, as well as the possibility of posttranscriptional modification, made amplifying the 3'UTRs highly difficult. All of these problems hindered our ability to clone our potential escapee 3'UTRs into our testing plasmid. Additionally, these problems hindered our ability to order synthetic 3'UTRs as the GC content interferes with the synthesis process. To address these problems, I turned to alternative methods in cloning. I used additives such as betaine and ethylene glycol, which are traditionally used to assist in overcoming translational stalling of the polymerase by disrupting GC interactions and weakening secondary structures. I tested temperature gradients to allow for better primer binding and to assist in the breaking down of secondary structures. I also attempted using alternative polymerases and redesigning primers. However, none of these process allowed for proper amplification of any of our targeted regions. The issue arises that the structural elements we are looking for as well as some posttranscriptional modifications in alignment with high GC content make an environment that disrupts polymerase activity and all around makes cloning difficult. Thus as was stated in the body of this work, genomic extraction, and the use of genomic DNA as the template for amplification seems to be the best solution for the cloning of 3'UTRs.

BIBLIOGRAPHY

D. Ganem, KSHV Infection and the Pathogenesis of Kaposi's Sarcoma. Annu. Rev. Pathol. Mech. Dis. 2006, 1:273-96 doi: 10.1146/annurev.pathol.1.110304.100133

E. Mesri, E. Cesarman, C. Boshoff, Kaposi's sarcoma herpesvirus/ Human herpesvirus-8 (KSHV/HHV8), and the oncogenesis of Kaposi's sarcoma. Nat Rev Cancer, 2010, 10(10) doi:10.1039/nrc2888

L. Yan, V. Majerciak, Z. Zheng, K. Lan, Towards Better Understanding of KSHV Life Cycl: from Transcription and Posttranscriptional Regulations to Pathogenesis. Virologica Sinica. 2019 doi: https://doi.org/10.1007/s12250-019-00114-3

K. Brulois, J. Jung, Interplay between Kaposi's sarcoma-associated herpesvirus and the innate immune system. Cytokine Growth Factor Rev. 2014, 25(5) doi:10.1016/j.cytogfr.2014.06.001.

R. Jenner, M. Albà, C. Boshoff, P. Kellam, Kaposi's Sarcoma-Associated Herpesvirus Latent and Lytic Gene Expression as Revealed by DNA Arrays. Journal of Virology. 2001, 75(2) DOI: 10.1128/JVI.75.2.891–902.2001

L. Burýšek, W. Yeow, B. Lubyová, M. Kellum, S. Schafer, Y. Huang, P. Pitha, Functional Analysis of Human Herpesvirus 8-Encoded Viral Interferon Regulatory Factor 1 and Its Association with Cellular Interferon Regulatory Factors and p300. Journal of Virology. 1999, 73(9). 0022-538X/99/$04.0010

F. Zhu, Y. Yuan, The ORF45 Protein of Kaposi's Sarcoma-Associated Herpesvirus Is Associated with Purified Virions. Journal of Virology. 2003, 77(7). DOI: 10.1128/JVI.77.7.4221–4230.2003

Y. Gwack, S. Hwang, H. Byun, C. Lim, J. Kim, E. Choi, J Choe. Kaposi's Sarcoma-Associated Herpesvirus Open Reading Frame 50 Represses p53-Induced Transcriptional Activity and Apoptosis. Journal of Virology. 2001, 75(13) DOI: 10.1128/JVI.75.13.6245–6248.2001

B. Johnston, E. Pringle, C. McCOrmick, KSHV activates unfolded protein response sensors but suppresses downstream transcriptional responses to support lytic replication. PLoS Pathogens. 2019, https://doi.org/10.1371/journal.ppat.1008185D

H. Zhang, G. Ni, B. Damania, ADAR1 Facilitates KSHV Lytic Reactivation by Modulating the RLR-Dependent Signaling Pathway. Cell Reports. 2020, 31(4) doi: 10.1016/j.celrep.2020.107564.

T. Tabtieng, A. Degterev, M. Gaglia, Caspase-Dependent Suppression of Type I Interferon Signaling Promtotes Kaposi's Sarcoma-Associated Herpesvirus Lytic Replication. Journal of Virology. 2018, 92(10). https://doi.org/10.1128/JVI.00078-18

K. Narayanan, S Makino, Interplay between viruses and host mRNA degradation. Biochimica et Biophysica Acta. 2013, 1829 http://dx.doi.org/10.1016/j.bbagrm.2012.12.003

H. Rivas, S. Schmaling, and M Gaglia, Shutoff of Host Gene Expression in Influenza A Virus and Herpesviruses: Similar Mechanisms and Common Themes. Viruses, 2016, 8(102) doi:10.3390/v8040102

W. Rodriguez, D. Macveigh-Fierro, J. Miles, M. Muller, Fated for decay: RNA elements targeted by viral endonucleases. Seminars in Cell & Developmental Biology. 2020. https://doi.org/10.1016/j.semcdb.2020.05.010

M. Gaglia, C. Rycroft, B. Glaunsinger, Transcriptome-Wide Cleavage Site Mapping on Cellular mRNAs Reveals Features Underlying Sequence-Specific Cleavage by the Viral Ribonuclease SOX. PLOS PATHOGENS, 2015. DOI:10.1371/journal.ppat.1005305

M. Gaglia, S. Covarrubias, W. Wong, B. Glaunsinger, A Common Strategy for Host RNA Degradation by Divergent Viruses. Journal of Virology, 2012, 86(17) DOI: 10.1128/JVI.01230-12

H. Lee, A. Patschull, C. Bagneris, H Ryan, C. Sanderson, B. Ebrahimi, I. Nobeli, T. Barret, KSHV SOX mediated host shutoff: the molecular mechanism underlying mRNA transcript processing. Nucleic Acids Research, 2017, 45(8). Doi: 10.1093/narlgkwl1340

S. Chandriani, D. Ganem, Host Transcript Accumulation during Lytic KSHV Infection Reveals Several Classes of Host Responses. PLoS One. 2007, 2(8) doi:10.1371/journal.pone.0000811

K. Clyde, B. Glaunsinger, Deep Sequencing Reveals Direct Targets of Gammaherpesvirus-Induced mRNA Decay and Suggests That Multiple Mechanisms Govern Cellular Transcript Escape. PLoS ONE, 2011, 6(5) doi:10.1371/journal.pone.0019655

M. Muller, B. Glaunsinger, Nuclease escape elements protect messenger RNA against cleavage by multiple viral endonucleases. PLoS Pathogens. 2017, https://doi.org/10.1371/journal.ppat.1006593

B. Glaunsinger, D. Ganem, Highly Selective Escape from KSHV-mediated Host mRNA Shutoff and Its Implications for Viral Pathogenesis. J Exp Med. 2004, 200(3) https://doi.org/10.1084/jem.20031881

S. Hutin, Y. Lee, B. Glaunsinger, An RNA Element in Human Interleukin 6 Confers Escape from Degradation by the Gammaherpesvirus SOX Protein. Journal of Virology, 2013, 87. doi:10.1128/JVI.00159-13

M. Muller, S. Hutin, O. Marigold, K. Li, A. Burlingame, B. Glaunsinger. A Ribonucleoprotein Complex Protects the Inerleukin-6 mRNA from Degradation by Distinct Herpesviral Endonucleases. PLoS Pathogens. 2015, DOI:10.1371/journal.ppat.1004899

W. Rodriguez, K. Srivastav, M. Muller, C19ORF66 Broadly Escapes Virus-Induced Endonuclease Cleavage and Restricts Kaposi's Sarcoma-Associated Herpesvirus. Journal of Virology. 2019, 93(12). DOI: 10.1128/JVI.00373-19

M. Chastain, I. Tinoco, Jr, Structural Elements in RNA. Progress in Nucleic Acid Research and Molecular Biology. 1991, 41. DOI: 10.1016/s0079-6603(08)60008-2

J. Doudna, E. Doherty, Emerging themes in RNA folding. Folding and Design. 1997, 2(5) https://doi.org/10.1016/S1359-0278(97)00035-7

R. Batey, R. Rambo, J. Doudna, Tertiary Motifs in RNA Structure and Folding. Angew Chem Int Ed Engl. 1999, 38(16) DOI: 10.1002/(sici)1521-3773(19990816)38:16<2326::aid-anie2326>3.0.co;2-3

S. Halder, D. Bhattacharyya, RNA Structure and dynamics: A base pairing. Progress in Biophysics and Molecular Biology. 2013, 113(2) https://doi.org/10.1016/j.pbiomolbio.2013.07.003

G. Varani, W. McClain, The G·U wobble base pair. EMBO reports. 2000, 1(1) doi: 10.1093/embo-reports/kvd001

P. Svoboda, A. Cara, Hairpin RNA: a secondary structure of primary importance. Cellular and Molecular Life Sciences. 2006, 63. DOI 10.1007/s00018-005-5558-5

J. Wyatt, J. Puglisi, I. Tinoco, Jr. RNA Folding: Pseudoknots, Loops, and Bulges. BioEssays. 1989, 11(4). DOI: 10.1002/bies.950110406

J. Wyatt, I. Tinoco, Jr. RNA Structural Elements and RNA Function. Cold Spring Harbor Press. The RNA World, Third Edition. 1993, Chapter 18.

T. Hermann, D. Patel. RNA bulges as architectural and recognition motifs. Structure. 2000, 8(3) https://doi.org/10.1016/S0969-2126(00)00110-6

L. Huang, D. Lilley. The Kink Turn, a Key Architectural Element in RNA Structure. Journal of Molecular Biology. 2016, 425(A) doi: 10.1016/j.jmb.2015.09.026

M. Chastain, I. Tinoco Jr., Structural Elements in RNA. Progress in Nucleic Acid Research and Molecular Biology. 1991, https://doi.org/10.1016/S0079-6603(08)60008-2

A. Lescoute, E. Westhof, Toplogy of Three-Way Junctions in Folded RNAs. RNA. 2006, 12(1). doi: 10.1261/rna.2208106.

E. Bindewald, R. Hayes, Y. Tingling, W. Kasprzakm B. Shapiro, RNAJunction: a database of RNA junctions and kissing loops for three-dimensional structural analysis and nanodesign. Nucleic Acid Research. 2008, 36 doi:10.1093/nar/gkm842

R. Batey, R. Rambo, J. Doudna, Tertiary Motifs in RNA Structure and Folding. Angew Chem Int Ed Engl. 1999, 38(16) DOI: 10.1002/(sici)1521-3773(19990816)38:16<2326::aid-anie2326>3.0.co;2-3

E. Westhof, P. Auffinger, RNA Tertiary Structure. Encyclopedia of Analytical Chemistry. www.els.net, Nature Publishing Group. 2001

M. Magnus, K. Kappel, R. Das, J. Bujnicki. RNA 3D structure prediction guided by independent folding of homologous sequences. BMC Bioinformatics, 2019, 20 https://doi.org/10.1186/s12859-019-3120-y

M. Abraham, O. Dror, R. Nussinov, H. Wolfson, Analysis and classification of RNA tertiary structures. RNA, 2008, 14(11) doi: 10.1261/rna.853208

D. Staple, S. Butcher, Pseudoknots: RNA Structures with Diverse Function. PLoS Biology, 2005, 3(6) DOI: 10.1371/journal.pbio.0030213

A. Steckleberg, Q. Vicens, J. Kieft, Exoribonuclease-Resistant RNAs Exist within both Coding and Noncoding Subgenomic RNAs. mBio. 2018, 9(6). https://doi.org/10.1128/mBio.02461-18

R. Ochsenreiter, I. Hofacker, M. Wolfinger, Functional RNA Structures in the 3'UTR of Tick-Borne, Insect-Specific and No-Known-Vector Flaviviruses. Viruses. 2019, 11(298) doi:10.3390/v11030298

S. Butcher, A. Pyle, The Molecular Interactions That Stabilize RNA Tertiary Structure: RNA Motifs, Patterns, and Networks. Accounts of Chemical Research, 2011, 44(11) https://doi.org/10.1021/ar200098t

C. Rossetto, G. Pari, PAN's Labyrinth: Molecular Biology of Kaposi's Sarcoma-Associated Herpesvirus (KSHV) PAN RNA, a Multifunctional Long Noncoding RNA. Viruses. 2014, 6(11) doi: 10.3390/v6114212

R. Mitton-Fry, S. DeGregorio, J. Wang, T. Steitz, J. Steitz, Poly(A) Tail Recognition by a Viral RNA Element Through Assembly of a Triple Helix. Science. 2010, 330 doi:10.1126/science.1195858.

F. Mignone, C. Gissi, S. Liuni, G. Pesole, Untranslated regions of mRNAs. Genome Biology, 2002, 3(3) doi: 10.1186/gb-2002-3-3-reviews0004

G. Pesole, F. Mignone, C. Gissi, G. Grillo, F. Licciulli, S. Liuni, Structural and functional features of eukaryotic mRNA untranslated regions. Gene. 2001, 276(1-2) https://doi.org/10.1016/S0378-1119(01)00674-6

K. Leppek, R. Das, M. Barna, Functional 5'UTR mRNA Structures in Eukaryotic Translation Regulation and How to Find Them. Nat Rev Mol Cell Biol. 2018, 19(3). DOI: 10.1038/nrm.2017.103

H. Liu, J. Yin, M. Xiao, C. Gao, A. Mason, Z. Zhao, Y. Lui, J. Li, D. Fu, Characterization and evolution of 5' and 3' untranslated regions in eukaryotes. Gene. 2012, 507 http://dx.doi.org/10.1016/j.gene.2012.07.034

J. Babendure, J. Babendure, J Ding, R. Tsien, Control of mammalian translation by mRNA structure near caps. RNA. 2006, 12 doi: 10.1261/rna.2309906

P. Araujo, K. Yoon, D. Ko, A. Smith, M. Qiao, U. Suresh, S. Burns, L. Penalva. Before It Gets Started: Regulating Translation at the 5' UTR. Comparative and Functional Genomics. 2012 doi:10.1155/2012/475731

C. Mayr, Regulation by 3'-Untranslated Regions. Annual Review of Genetics. 2017, 13(29) https://doi.org/10.1146/annurev-genet-120116-024704

S. Lianoglou, V. Garg, J. Yang, C. Leslie, C. Mayr, Ubiquitously transcribed genes use alternative polyadenylation to achieve tissue-specific expression. Genes & Development. 2013, 27(21) doi: 10.1101/gad.229328.113

C. Mayr, What are 3' UTRs Doing?. Cold Spring Harb Perspect Biol. 2019, 11(10) doi: 10.1101/cshperspect.a034728

Eva Matoulkova, Eva Michalova, Borivoj Vojtesek & Roman Hrstka, The role of the 3' untranslated region in post-transcriptional regulation of protein expression in mammalian cells., RNA Biology, 2012, 9(5) DOI: 10.4161/rna.20231

C. Chen, A. Shyu, AU-Rich elements: characterization and importance in mRNA degradation. Trends Biochem Sci. 1995, 20(11) DOI: 10.1016/s0968-0004(00)89102-1

G. Shaw, R. Kamen, A conserved AU sequence from the 3’ untranslated region of GM-CSF mRNA mediates selective mRNA degradation. Cell. 1986, 46(5) doi: 10.1016/0092-8674(86)90341-7.

D. Caput, B. Beutler, K. Hartog, R. Thayer, S. Brown-Shimer, A. Cerami, Identification of a Common Nucleotide Sequence in the 3’untranslated Region of mRNA Molecules Specifying Inflammatory Mediators. Proc Natl Acad Sci USA. 1986, 83(6) doi: 10.1073/pnas.83.6.1670.

C. Barreau, L. Paillard, H. Osborne, AU-rich elements and associated factors: are there unifying principles?. Nucleic Acids Research. 2005, 33(22) doi:10.1093/nar/gki1012

H. Otsuka, A. Fukao, Y. Funakami, K. Duncan, T. Fujiwara, Emerging Evidence of Translational Control by AU-Rich Element-Binding Proteins. Frontiers in Genetics. 2019, 10(332) doi: 10.3389/fgene.2019.00332

A. Halees, E. Hitti, M. Al-Saif, L. Mahmoud, I. Vlasova-St. Louis, D. Beisang. P. Bohjanen, L. Khabar, Global assessment of GU-rich regulatory content and function in the human transcriptome. RNA Biology. 2011, 8(4) DOI: 10.4161/rna.8.4.16283

I. Vlasoca-St. Lous, P BOhjanen, Coordinate Regulation of mRNA Decay Networks by GU-rich Elements and CELF1. Curr Opin Genet Dev. 2011, 21(4) doi:10.1016/j.gde.2011.03.002

E. Hollams, K. Giles, A. Thomson, P. Leedman, mRNA Stability and the Control of Gene Expression: Implications for Human Disease. Neurochemical Research. 2002, 27(10) DOI: 10.1023/A:1020992418511

R. Erlitzki, J. Long, E. Theil, Multiple, Conserved Iron-responsive Elements in the 3'Untranslated Region of Transferrin Receptor mRNA Enhance Binding of Iron Regulatory Protein 2. Journal of Biological Chemistry. 2002, 277(45) doi: 10.1074/jbc.M207918200

K. Korotkov, S. Novoselov, D. Hatfield, V. Gladshev, Mammalian Selenoprotein in Which Selenocysteine (Sec) Incorporation is Supported by a New Form of Sec Insertion Sequence Element. Molecular and Cellular Biology. 2002, 22(5) DOI: 10.1128/MCB.22.5.1402–1411.2002

H. Mix, A. Lobanov, V. Gladyshev, SECIS elements in the coding regions of selenoprotein transcripts are functional in higher eukaryotes. Nucleic Acids Research. 2007, 35(2) doi:10.1093/nar/gkl1060

M. Hentze, A. Castello, T. Schwarzi, T. Preiss, A brave new world of RNA-binding proteins. Nature Reviews Molecular Cell Biology. 2018, 19 doi:10.1038/nrm.2017.130

J. Spiegel, S. Adhikari, S. Balasubramanian, The Structure and Function of DNA G-Quadruplexes. Trends in Chemistry. 2019, 2(2) https://doi.org/10.1016/j.trechm.2019.07.002

S. Covarrubias, J. Richner, K. Clyde, Y. Lee, B. Glaunsinger, Host Shutoff Is a Conserved Phenotype of Gammaherpesvirus Infection and Is Orchestrated Exclusively from the Cytoplasm. Journal of Virology. 2009, 83(18) doi:10.1128/JVI.01051-09

M. Buisson, T. Géoui, D. Flot, N. Tarbouriech, M. Ressing, E. Wiertz, W. Burmeister, A Bridge Crosses the Active-Site Canyon of the Epstein-Barr Virus Nuclease with DNase and RNase Activities. J. Mol Biol. 2009, 391 doi:10.1016/j.jmb.2009.06.034

J. Kosinski, M. Feder, J. Bujnicki, The PD-(D/E)XK superfamily revisited: identification of new members among proteins involved in DNA metabolism and functional predictions for domains of (hitherto) unknown function. BMC Bioinformatics. 2005, 12(6) doi: 10.1186/1471-2105-6-172.

S. Covarrubias, M. Gaglia, G. Kumar, W. Wong, A. Jackson, B. Glaunsinger, Coordinated Destruction of Cellular Messages in Translation Complexes by the Gammaherpesvirus Host Shutoff Factor and the Mammalian Exonuclease XRN1. PLoS Pathogens. 2011, 7(10) doi:10.1371/journal.ppat.1002339

E. Abernathy, S. Gilbertson, R. Alla, B. Glaunsinger, Viral Nucleases Induce an mRNA Degradation-Transcription Feedback Loop in Mammalian Cells. Cell Host Microbe. 2015, 18(2) doi: 10.1016/j.chom.2015.06.019.

E. Abernathy, K. Clyde, R. Yeasmin, L. Krug, A. Burlingame, L. Coscoy, B. Glaunsinger. Gammaherpesviral Gene Expression and Virion Composition Are Broadly Controlled by Accelerated mRNA Degradation. PLoS Pathogens. 2014, https://doi.org/10.1371/journal.ppat.1003882

A. Mendez, C. Vogt, J. Bohne, B. Glaunsinger. Site specific target binding controls RNA cleavage efficiency by the Kaposi's sarcoma-associated herpesvirus endonuclease SOX. Nucleic Acids Research. 2018, 46(22) doi: 10.1093/nar/gky932

M. Somla, G. Rice, S. Busan, N. Siegfried, K. Weeks, Selective 2'-hydroxyl acylation analyzed by primer extension and mutational profiling (SHAPE-MaP) for direct, versatile and accurate RNA structural analysis. Nature Protocols. 2015, 10(11) doi:10.1038/nprot.2015.103

S. Martin, C. Blankenship, J. Rausch, J. Sztuba-Solinska, Using SHAPE-MaP to probe small molecule-RNA interactions. Methods. 2019, 167(1) https://doi.org/10.1016/j.ymeth.2019.04.009

S. Busan, C. Weidmann, A. Sengupta, K. Weeks, Guidelines for SHAPE Reagent Choice and Detection Strategy for RNA Structure Probing Studies. Biochemistry. 2019, 58 DOI: 10.1021/acs.biochem.8b01218